

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 1, 2002, 11:18:53 ; Search time 32.5 Seconds
(without alignments)
1233.774 Million cell updates/sec

Title: US-09-839-894-10

Perfect score: 1886
Sequence: 1 MNKILFIFTLFFSSVLFTHFA.....EAGQYMGNIWTFPPSSQTL 361

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: A.GeneSeq.032802.*
2: /SIDSI/gcgdata/hold-geneSeq/geneSeq-emb1/AA1980.DAT.*
3: /SIDSI/gcgdata/hold-geneSeq/geneSeq-emb1/AA1981.DAT.*
4: /SIDSI/gcgdata/hold-geneSeq/geneSeq-emb1/AA1982.DAT.*
5: /SIDSI/gcgdata/hold-geneSeq/geneSeq-emb1/AA1983.DAT.*
6: /SIDSI/gcgdata/hold-geneSeq/geneSeq-emb1/AA1984.DAT.*
7: /SIDSI/gcgdata/hold-geneSeq/geneSeq-emb1/AA1985.DAT.*
8: /SIDSI/gcgdata/hold-geneSeq/geneSeq-emb1/AA1986.DAT.*
9: /SIDSI/gcgdata/hold-geneSeq/geneSeq-emb1/AA1987.DAT.*
10: /SIDSI/gcgdata/hold-geneSeq/geneSeq-emb1/AA1988.DAT.*
11: /SIDSI/gcgdata/hold-geneSeq/geneSeq-emb1/AA1989.DAT.*
12: /SIDSI/gcgdata/hold-geneSeq/geneSeq-emb1/AA1990.DAT.*
13: /SIDSI/gcgdata/hold-geneSeq/geneSeq-emb1/AA1991.DAT.*
14: /SIDSI/gcgdata/hold-geneSeq/geneSeq-emb1/AA1992.DAT.*
15: /SIDSI/gcgdata/hold-geneSeq/geneSeq-emb1/AA1993.DAT.*
16: /SIDSI/gcgdata/hold-geneSeq/geneSeq-emb1/AA1994.DAT.*
17: /SIDSI/gcgdata/hold-geneSeq/geneSeq-emb1/AA1995.DAT.*
18: /SIDSI/gcgdata/hold-geneSeq/geneSeq-emb1/AA1996.DAT.*
19: /SIDSI/gcgdata/hold-geneSeq/geneSeq-emb1/AA1997.DAT.*
20: /SIDSI/gcgdata/hold-geneSeq/geneSeq-emb1/AA1998.DAT.*
21: /SIDSI/gcgdata/hold-geneSeq/geneSeq-emb1/AA1999.DAT.*
22: /SIDSI/gcgdata/hold-geneSeq/geneSeq-emb1/AA2000.DAT.*
23: /SIDSI/gcgdata/hold-geneSeq/geneSeq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	927.5	49.2	364	20	AAV22326	Pilin protein CotD
2	275.5	14.6	359	22	AAV45919	S. enterica serova
3	119.5	6.3	770	21	AAV50814	D. melanogaster ac
4	114	6.0	650	15	AAV47575	Alpha-2-glycyltulin o
5	111.5	5.9	1005	21	AAV01833	Haemophilus influe
6	111.5	5.9	1011	21	AAV01832	Haemophilus influe
7	110.5	5.9	1529	14	AAV41732	High molecular wei
8	110.5	5.9	1601	18	AAV30292	Non-typeable Haemo
9	109	5.8	2383	21	AAV15945	E. coli proliferat
10	107	5.7	1095	21	AAV01835	Haemophilus influe
11	107	5.7	1101	21	AAV01834	Haemophilus influe

12	106	5.6	671	17	AAV85290	Streptococcus faec
13	106	5.6	1222	21	AAV01830	H. influenzae stra
14	106	5.6	1228	21	AAV01828	Haemophilus influe
15	101	5.4	1338	14	AAV41731	High molecular wei
16	101	5.4	1598	18	AAV30291	Non-typeable Haemo
17	100.5	5.3	1004	21	AAV01841	Haemophilus influe
18	100.5	5.3	1010	21	AAV01840	Haemophilus influe
19	99.5	5.3	969	21	AAV01827	Haemophilus influe
20	99.5	5.3	975	21	AAV01826	Haemophilus influe
21	99	5.2	992	21	AAV01843	Haemophilus influe
22	99	5.2	998	21	AAV01842	Haemophilus influe
23	97	5.1	418	21	AAV30907	Arabidopsis thaila
24	97	5.1	454	21	AAV30906	Arabidopsis thaila
25	97	5.1	957	21	AAV01839	Haemophilus influe
26	97	5.1	963	21	AAV01838	Haemophilus influe
27	96	5.1	430	21	AAV37170	Arabidopsis thaila
28	96	5.1	513	21	AAV39277	Arabidopsis thaila
29	96	5.1	631	21	AAV39276	Arabidopsis thaila
30	96	5.1	683	21	AAV39275	Arabidopsis thaila
31	95.5	5.1	825	21	AAV25547	Eucalyptus grandis
32	94.5	5.0	567	20	AAV37170	Amino acid sequenc
33	94.5	5.0	665	22	AAV68244	Drosophila melanog
34	94.5	5.0	2353	17	AAV93933	Haemophilus adhesi
35	94.5	5.0	2411	21	AAV23860	Haemophilus influe
36	94	5.0	394	21	AAV37668	Arabidopsis thaila
37	93.5	5.0	715	20	AAV05625	HIV-1 group O isol
38	93.5	5.0	1464	22	AAV71111	Drosophila melanog
39	93	4.9	354	18	AAV55510	H. pylori ORF 04ge
40	93	4.9	444	18	AAV55696	H. pylori ORF 13ae
41	93	4.9	444	20	AAV17193	H. pylori outer me
42	93	4.9	808	22	AAV58477	Drosophila melanog
43	93	4.9	1167	18	AAV31504	Nematoe toxin 167
44	93	4.9	1167	18	AAV10653	Bacillus thuringie
45	92.5	4.9	1073	21	AAV01837	Haemophilus influe

ALIGNMENTS

RESULT 1	AAV22326	standard; Protein: 364 AA.
ID	AAV22326	
XX	AAV22326	
AC	22-SEP-1999	(first entry)
XX		
DE	Pilin protein CotD.	
XX		
KW	CS2 gene cluster; CotA; CotB; CotC; CotD; pilin protein; immunogen;	
KW	enterotoxigenic E. coli; human upper intestine; diarrhoeal disease;	
KW	enteric infection; therapy.	
XX		
OS	Escherichia coli.	
XX		
PN	US5932715-A.	
XX		
PD	03-AUG-1999.	
XX		
PF	07-JUN-1995; 95US-0483101.	
XX		
PR	07-JUN-1995; 95US-0483101.	
XX		
PA	(UYEM-) UNIV EMORY.	
XX		
PI	Caron J, Froehlich B, Scott JR.	
XX		
DR	WPI, 1999-443623/37.	
XX	N-PSDB; AAV84848.	
PT	Isolated nucleic acids encoding Escherichia coli CS2 pilin proteins	
PT	useful for vaccinating against diarrhoeal diseases caused by	
PT	Escherichia coli	

XX Claim 3; Column 45-48; 35pp; English.
PS This sequence represents a CS2 pilin protein of the invention, encoded by
XX (which also encodes Cota, CotaC, and CotaB). CS2 pilins are long
CC proteinaceous molecules thought to mediate attachment of enterotoxigenic
CC E. coli (ETEC) to and/or promote colonisation of the human upper
CC intestine. The CS2 gene cluster may be used to produce immunogens for
CC vaccinating patients against diarrhoeal diseases caused by ETEC bacteria.
CC This type of enteric infection is a major cause of death among infants in
CC developing countries and in immunocompromised (e.g. Acquired Immune
CC Deficiency Syndrome (AIDS) or elderly adults. The vaccine comprises more
CC than 1 antigenic determinant (epitopes) from more than 1 pilus type to be
CC effective against more than 1 type of ETEC infection.
XX
XX Sequence 364 AA:
SO
Query Match 49.2%; Score 927.5; DB 20; Length 364;
Best Local Similarity 48.1%; Pred. No. 7.5e-79;
Matches 176; Conservative 65; Mismatches 118; Indels 7; Gaps 3;
QY 1 MNILFLTFEFSVLTFAVSADKIPGDSINIRPRORN--ESSPKNINLNNITAY 58
DB 1 LKVLITVLMFLCSQVYQGWCHNVEAGSINKTESIPIDIRSAASYPAHYIFHEHAGY 60
QY 59 SESHTLYDRMTFLCLSHNTLNGACPTSENSSVSGENITITQTEKRLIKRELQIK 118
DB 61 NKAHSLIDRMTFLCMSSTDSKGAQCPGNSKSS--QGENIKLITFEKSLARKLNLK 118
QY 119 GYKQLLEKSVNC--PSGLTLNSAHFNCKNNAASGASLYLIPAGELKNLPFGGIMDATL 175
DB 119 GYKRLFYESDRCLHYDCKMLNLSHTVKCVSFCFQVDFLYIPGGLDGLTGIWEATL 178
QY 176 KLVKRRRYSTYTYTINITIKLTDKGNIOIWLPOKRSARVDLNLPRPGCGYTGNSV 235
DB 179 ELTVKCHYDNLKYVNLTVDLTDKGNLQVWPKFNSDPRIDNLIRPEGNGKYSGSNVL 238
QY 236 DMCFYGVGYSTNSSLERFODNNPKSDGKFLRKINDDTKEIAYTSLLAGSLTPNTG 295
DB 239 EMCLYGGYSHSGSIEMTFQDASQGNENEYLIKTEPIKLIYKLSIIIGRETFPNAG 298
QY 296 TSLINADASLETNNRITAVTPEISVPLCPWGRLOLDKAVENPEAGQYMGINIVTF 355
DB 299 EATLINDTSLIFINMIRIKESVLPQISIPVLCWPAHLTFMSELNPPPEAGYSGILNVTF 358
QY 356 PSSQTL 361
DB 359 PSSSSL 364

RESULT 2
AAB45919
ID AAB45919 standard; Protein; 359 AA.
XX
XX AAB45919;
AC
XX
XX 23-MAR-2001 (first entry)
DT
XX
XX S. enterica serovar Typhi tcfD fimbrial subunit protein.
DE
XX
XX Fimbrial protein; saf; tcf; vaccine; gene therapy; immunization;
KW tcf insert; detection.
XX
XX Salmonella typhi.
OS
XX
XX WO200073336-A1.
PN
XX
XX 07-DEC-2000.
PD
XX
XX 26-MAY-2000; 2000WO-SE01079.
PF
XX
XX 28-MAY-1999; 99SE-0001961.
PR

XX
PA (ACTI-) ACTIVE BIOTECH AB.
XX
XX Polkesson A, Normark S, Loeffdahl S;
PI
XX
XX WPI; 2001-061512/07.
DR
XX
XX N-PSDB; AAC82926.
DR
XX
XX Fimbriae proteins of Salmonella enterica subspecies I bacteria, useful
PT for producing vaccines against the bacterial subspecies and for
PT detecting the bacteria
XX
XX Disclosure; Page 72-73; 77pp; English.
XX
XX This invention describes the novel proteins (saf and tcf) (I) encoded by
CC a DNA sequence of a gene encoding the precursor of the saf fimbriae unit
CC of Salmonella enterica subspecies I or a DNA sequence of the gene
CC encoding the tcf fimbriae unit of S. enterica subspecies I serovar Typhi
CC The products of the invention can be used as vaccines or for gene
CC therapy. Such vaccines are useful for protection against diseases caused
CC by S. enterica subspecies I or S. enterica subspecies I serovar Typhi.
CC The saf and tcf proteins from S. enterica subspecies I bacteria are
CC useful for active or passive immunization in mammals. The nucleotide
CC sequences are useful for constructing vectors for use as vaccines for
CC insertion into attenuated bacteria in constructing a recombinant viral
CC vaccine, or for direct inoculation of a nucleic acid vaccine. The protein
CC or antigenic fragments, nucleic acid sequences, and antibodies are useful
CC in molecular diagnostic assays for the detection of S. enterica
CC subspecies I.
XX
XX Sequence 359 AA:
SO
Query Match 14.6%; Score 275.5; DB 22; Length 359;
Best Local Similarity 29.7%; Pred. No. 2.1e-17;
Matches 90; Conservative 48; Mismatches 130; Indels 35; Gaps 11;
QY 70 FLCISHTNLNGACPTSENSSVSGENITITQTEKRLIKRELQIKGKQLKFSVN 129
DB 79 WVCISHTNNEGACEETHLYWYAFGAYSKIRIRFEGISHAEITL-----ILIGSVR 131
QY 130 --CPSGTLTNSAHFNCKNNAAS--GASLYLIPAGELKNLPFGGIMDATLKL-RVKRRY 183
DB 132 DACYTG-----INMAAAGCWGRSLKLRIPSELAKIPSGCWKALVIDYQWGG 183
QY 184 SETYGYTITITIKLND--KGNIOIWLPOKRSARVDLNLPRPGCGYTGNSVMCRY 240
DB 184 DDPJGFTLIDTLNVCHFAENAAIYFPQFGTATPRVDNLHNMASQMSGRANLDMCLY 243
QY 241 DGSTNSSLERFODNNPKSDGKFLRKINDDTKEIAYTSLLAGSLTPNTGSLNT 300
DB 244 DG-GVKARSLGKLEGSNGSGTFGYIKSDASAT--IDYAVSMYGRISPRVYRGEFSL 300
QY 301 ADAASLETNNRITAVTPEISVPLCPWGRLOLDK---VENPEAGQYMGINIVTFPS 357
DB 301 DNVDKAATR-----PVLPQGRQVRCVPPPLTTCQPFNIREKRSGEYGGTLTVMLNG 355
QY 358 SQT 360
DB 356 TGT 358

RESULT 3
AA50814
ID AA50814 standard; Protein; 770 AA.
XX
XX AA50814;
AC
XX
XX 17-FEB-2000 (first entry)
DT
XX
XX D. melanogaster acetyl-choline receptor protein from clone Da7.
DE
XX
XX Acetyl-choline receptor; nicotinic; insect; insecticide; screening;
KW

Best Local Similarity 20.0%; Pred. No. 1.8;
Matches 78; Conservative 65; Mismatches 157; Indels 90; Gaps 15;

```
OY 13 SSVLFPAVNSADKIPGDESDITNIFGPRDRNESSPKHNLNHNITAY----SESHLYDRMT 69
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 762 skiacllasanngvlnaenaantvsnvadegs---npindhvltfavlsagsatfngnq 818
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 70 FCLLSHNTLNGACPTSENPSSSVSGENITITLQFTEKSLI-----K 112
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 819 ----taktvngla-tfdlkskqdehvtvengvkgqllvsfygdsstgavdlqsk 873
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 113 RELQIKGYKOL-LFKSVNCPSPGLTTLNSAHFNCNKNAASGASLYLIPAGELKMLPGGIW 171
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 874 neevadgndsvlmtatvrakgnllndvmtvfnvnsaeklsqtenvnsd-----gla 926
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 172 DATLKIRKRRRISEYGYTTITITIKLTDKNIQIWLPOPKSDARYDLNIRPGSGTYIG 231
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 927 tatlt-----slkngdyfvltasvsgsqanqvffigdstaaltlsv-psgdltv-- 976
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 232 RNSVDMCFDGYSTNSS-----SLEIRFQDNNPKSDGKFLRKINDPTKEIAYTLSDLLAG 287
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 977 -----tntlapgymatclqdkngnplkdkeltfsvpnd---vaskfslsngg 1019
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 288 KSLTPTNGTSL-----NIADAASLETMMNRITAYTMPETISVPVLC 327
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 1020 kqmtsdngvafasltgltgltmmlarlnansvnsdqpmntfvadkdravvvlqtskaell 1079
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 328 WPG--RLQIDAKVENPEACGYMGNINVTFT 355
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 1080 gngvdehtlatvkd-pnhpvaqlvntft 1108
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
```

RESULT 10

```
AAB01835
ID AAB01835 standard; Protein; 1095 AA.
AC AAB01835;
XX
XX 11-SEP-2000 (first entry)
DE Haemophilus influenzae strain LCDC2 mature HMW1A protein, SEQ ID NO:45.
XX
XX Mature HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KM non-tyeable Haemophilus influenzae; NTNH1; non-encapsulated;
KM recombinant production; Escherichia coli; antibacterial; vaccine;
KM human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KM detection; diagnosis.
XX
XX Haemophilus influenzae strain LCDC2.
XX
XX WO200020609-A2.
XX
XX 13-APR-2000.
XX
XX 07-OCT-1999; 99WO-CA00938.
XX
XX 07-OCT-1998; 98US-0167568.
XX
XX 08-DEC-1998; 98US-0206942.
XX
XX (CONN-) CONNAUGHT LAB LTD.
XX
XX Loosmore SM, Yang Y, Klein MH;
XX
XX WPI; 2000-303789/26.
XX
XX N-PSDB; AAA52184.
XX
XX Nucleic acid molecule for producing recombinant high molecular weight
PT proteins of Haemophilus which are used as a vaccine to provide
PT protection against Haemophilus induced diseases in humans -
XX
XX Claim 8; Fig 22A-P; 307pp; English.
XX
XX The invention relates to the recombinant production of Haemophilus
CC
```

```
CC influenzae high molecular weight (HMW) proteins in Escherichia coli. The
CC expression construct used to effect recombinant expression comprises a
CC promoter functional in E. coli (e.g., the T7 promoter) operably linked
CC to a modified hmwABC operon from a non-tyeable (non-encapsulated) H.
CC influenzae (NTNH1). Most HMW-expressing NTNH1 strains contain two hmw gene
CC clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,
CC hmwB and hmwC genes. The hmwA genes encode the structural HMW proteins
CC and the hmwB and hmwC genes encode accessory proteins which are
CC responsible for post-translational processing and secretion of the HMW
CC proteins. The modified hmwABC operon used in the expression construct of
CC the invention contains an A gene modified such that it encodes only the
CC mature HMW. The invention also discloses hmwA genes (AAA52175-A52198)
CC and HMW proteins (AAB01824-A01849) from the non-tyeable H. influenzae
CC strains Joyce, KL, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and
CC vectors are used for the production of recombinant H. influenzae HMW
CC proteins which can be used as vaccines to mediate a humoral or
CC cell-mediated immune response to provide protection against diseases in
CC humans caused by H. influenzae (e.g., otitis media, epiglottitis,
CC pneumonia and tracheobronchitis). The HMW proteins are also useful as
CC antigens in immunoassays for detecting antibodies against Haemophilus,
CC HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
CC HMW proteins can be used to isolate and clone hmw genes from other
CC non-tyeable strains of Haemophilus via hybridisation reactions. The
CC present sequence represents a mature HMW protein from a non-tyeable
CC strain of H. influenzae.
XX
XX Sequence 1095 AA:
```

Query Match 5.7%; Score 107; DB 21; Length 1095;
Best Local Similarity 24.6%; Pred. No. 0.86;
Matches 71; Conservative 44; Mismatches 110; Indels 64; Gaps 15;

```
OY 45 SPRKHNILNHNITAYSESHLYD-----RMFPLCLSSHNTLNGACPTSENPSSSS 93
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 281 slkplvsn---vhqgnhtlfnqnvsvlggdyvfnfhassnwhchgyvlxskntnaae 337
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 94 VSGENITITLQFTEKSLIRKINDPTKEIAYTLSDLLAG 327
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 338 gs-----slrf-----ksegstrlaf-----tledltlnaagnislngqldg 378
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 154 YIYIPAGELKMLPGGIMPATLKLKRRRYSEYGYTYI-----NIRIKLTDKNIQIWP 209
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 379 nlqkslvaknltfeg---gnltlaadkpkiekgiltvkeganvltreanygndksal- 434
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 210 QFKSDARVDLNRPTGGTYIGRN-----SDVDCFYDGYSTNSSSLEIRPDNNPKSD-- 262
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 435 slrgnvtknlgltvtgsalnleknltvegsakflnapnysfnvsgl-----fdngksnls 490
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 263 ---GKFYLRKINDPTKEIAYTLSDLLAGKSLTPTNGSLNIAD 302
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 491 lakgahfkdn-nlkslnltltnsdasyrtlileg-nltinsng-dlnld 536
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
```

RESULT 11

```
AAB01834
ID AAB01834 standard; Protein; 1101 AA.
AC AAB01834;
XX
XX 11-SEP-2000 (first entry)
DE Haemophilus influenzae strain LCDC2 HMW1A protein, SEQ ID NO:43.
XX
XX
XX Haemophilus influenzae strain LCDC2 HMW1A protein, SEQ ID NO:43.
XX
XX HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KM non-tyeable Haemophilus influenzae; NTNH1; non-encapsulated;
KM recombinant production; Escherichia coli; antibacterial; vaccine;
KM human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KM detection; diagnosis.
XX
XX Haemophilus influenzae strain LCDC2.
XX
XX WO200020609-A2.
XX
XX
```

XX 13-APR-2000.
PD
XX
PE 07-OCT-1999; 99WO-CA00938.
XX
PR 07-OCT-1998; 98US-0167568.
PR 08-DEC-1998; 98US-0206942.
XX
PA (CONN-) CONNBUGHT LAB LTD.
XX
PI
XX Loosmore SM, Yang Y, Klein MH;
PI
XX WPI: 2000-303789/26.
DR
DR N-PSDB: AAA52183.
XX
XX Nucleic acid molecule for producing recombinant high molecular weight
PT proteins of Haemophilus which are used as a vaccine to provide
PT protection against Haemophilus induced diseases in humans -
XX
XX
XX Claim 12; Fig 22A-P; 307pp; English.

The invention relates to the recombinant production of Haemophilus influenzae high molecular weight (HMW) proteins in Escherichia coli. The expression construct used to effect recombinant expression comprises a promoter functional in E. coli (e.g., the T7 promoter) operably linked to a modified hmwaBC operon from a non-typable (non-encapsulated) H. influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene clusters termed hmwABC and hmw2ABC. Each hmwaBC operon comprises hmwa, hmwB and hmwC genes. The hmwa genes encode the structural HMWA proteins and the hmwB and hmwC genes encode accessory proteins which are responsible for post-translational processing and secretion of the HMWA proteins. The modified hmwaBC operon used in the expression construct of the invention contains an A gene modified such that it encodes only the mature HMWA. The invention also discloses hmwa genes (AAA52117-A52198) and HMWA proteins (AAB01824-B01849) from the non-typable H. influenzae strains Jovy, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and vectors are used for the production of recombinant H. influenzae HMW proteins which can be used as vaccines to mediate a humoral or cell-mediated immune response to provide protection against diseases in humans caused by H. influenzae (e.g., otitis media, epiglottitis, pneumonia and tracheobronchitis). The HMW proteins are also useful as antigens in immunoassays for detecting antibodies against Haemophilus, HMW proteins and/or HMW peptides. The nucleotide sequences encoding the HMW proteins can be used to isolate and clone hmw genes from other non-typable strains of Haemophilus via hybridisation reactions. The present sequence represents an HMWA protein from a non-typable strain of H. influenzae.

SQ Sequence 1101 AA;

Query Match	5.7%	Score 107;	DB 21;	Length 1101;
-------------	------	------------	--------	--------------

Matches 71; Conservative 44; Mismatches 110; Indels 64; Gaps 15;

45 SPKHNLNNHITAYSESHLYD-----RMTFLCLSSHNTLNGACPTSENPSSS

Db 287 slkppivsn---vhdgntlfngvsvlgygdvnfhfnasssnhwlthgvlksgnfnaas

QY 94 VSGETNITLOFTEKRSLLIKRELOIKGYKQLLFKSVNCPSGLTLSAHCNCNKAASGAS

Db 344 gs-----slr-----ksegstrtat-----tiesdltlnatggnlslnqvagld

134 LLI PAGE LK NLFPGIWDATLKLVRKRRYSSEIYGYTL-----NLYIKLIDKGNIQIWL

DD 365 mlgksivanknrlcleg--gnlllaadkkrprieikgnltcvkeganvlllsanhygnaksal

[illegible]

THE UNIVERSITY OF CHICAGO LIBRARY

[illegible]

Db	497	jakgahhkldin-nkxlnitnnsayrtllg-nltnng-dlnld	542
	RESULT	12	
	AA85290		
ID	AA85290	standard; Protein; 671 AA.	
XX			
AC	AA85290;		
XX			
DT	02-APR-1996	(first entry)	
XX			
DE	Streptococcus faecalis autolysin.		
XX			
KW	Lysin; autolysis; culture; lactic acid bacteria; fermentation;		
KW	cheese; foodstuffs; induction.		
XX			
OS	Streptococcus faecalis.		
XX			
PN	W09531561-A1.		
XX			
PD	23-NOV-1995.		
XX			
PF	12-MAY-1995;	95WO-NL00170.	
XX			
PR	12-MAY-1994;	94EP-0201353.	
XX			
PA	(UNIL) QUEST INT BV.		
XX			
PI	Buist G, Kok J, Ledebor AM, Venema G;		
XX			
DR	WPI; 1996-010946/01.		
XX			
PT	Lysis of a culture of lactic acid bacteria in, e.g. cheese		
PT	production - by in situ prodn. of an auto:lysin, regulated by an		
PT	inducible promoter.		
XX			
PS	Disclosure; Page 66-69; 103pp; English.		
XX			
CC	In situ production of a homologous autolysin or a heterologous		
CC	autolysin from a food grade Gram positive bacteria, can be used in		
CC	a process for the lysis of a culture of lactic acid bacteria. The		
CC	process can be used in the manufacture of products containing		
CC	cultures of lactic acid bacteria e.g. cheese, where the culture is		
CC	lysed following the completion of fermentation. The enhanced		
CC	induction of the autolysin is performed some hours after the		
CC	fermentation is finished. No extra lysin needs to be added and the		
CC	lysin does not need to be isolated or encapsulated. The time of		
CC	lysis can be precisely controlled. This is the Streptococcus		
CC	faecalis autolysin.		
XX			
SO	Sequence	671 AA;	

Query Match	5.6%	Score 106;	DB 17;	Length 671;
Best Local Similarity	23.6%	Pred. No. 0	51;	
Matches	73;	Conservative	39;	Mismatches 103;
				Indels 94;
				Gaps 17;

QY 41 RNESSPKHNI-LNNHITAYSESHLYDRMTFLCLSSHNTLNGACPTSENPSSSSV----- 94

Db 309 ryatdpsynaklINVltay--nltqydtprssgntlggtvnpptggsnngsgnttytvc 366

95 SGET--NITLQF---TEKRSL--IKRELQIKGYKQLLFKSVNCPSSGLTNSAHFNCNKN 146

Db 367 sgdtlnkiaqygvsvanlrswnglsgdllrvgqklivkkga---sgntggsgnggsnnn 423

14 / AASGASLYLIPAGEELKNLEPGIWDATLKLKRVKRRISLIGITINILINKLIDKNGIQL 208

424 -qsyulrYcVksyucln-----kIaaqYg-----vLv-----aullS 425

[illegible][illegible]

PF 07-OCT-1999; 99WO-CA00938.
 XX 07-OCT-1998; 98US-0167568.
 PR 08-DEC-1998; 98US-0206942.
 XX (CONN-) CONNAUGHT LAB LTD.
 XX Loosmore SM, Yang Y, Klein MH;
 XX WPI: 2000-303789/26.
 DR N-PSDB; AAA52179.
 XX
 PT Nucleic acid molecule for producing recombinant high molecular weight
 PT proteins of Haemophilus which are used as a vaccine to provide
 PT protection against Haemophilus induced diseases in humans -
 XX
 PS Claim 12; Fig 20A-R; 307pp; English.

CC The invention relates to the recombinant production of Haemophilus
 CC Influenzae high molecular weight (HMW) proteins in Escherichia coli. The
 CC expression construct used to effect recombinant expression comprises a
 CC promoter functional in E. coli (e.g., the T7 promoter) operably linked
 CC to a modified hmwaB operon from a non-typeable (non-encapsulated) H.
 CC Influenzae (NTH1). Most HMW-expressing NTH1 strains contain two hmw gene
 CC clusters termed hmw1ABC and hmw2ABC. Each hmwaB operon comprises hmwaB,
 CC hmwaC and hmwaD genes. The hmwa genes encode the structural HMWA proteins
 CC and the hmwaB and hmwaC genes encode accessory proteins which are
 CC responsible for post-translational processing and secretion of the HMWA
 CC proteins. The modified hmwaB operon used in the expression construct of
 CC the invention contains an A gene modified such that it encodes only the
 CC mature HMWA. The invention also discloses hmwa genes (AAA52175-A52198)
 CC and HMWA proteins (AAB01824-B01849) from the non-typeable H. Influenzae
 CC strains Joyce, K1, K21, LCC02, PMH1, 15 and 12. The nucleic acids and
 CC vectors are used for the production of recombinant H. Influenzae HMW
 CC proteins which can be used as vaccines to mediate a humoral or
 CC cell-mediated immune response to provide protection against diseases in
 CC humans caused by H. Influenzae (e.g., otitis media, epiglottitis,
 CC pneumonia and tracheobronchitis). The HMW proteins are also useful as
 CC antigens in immunoassays for detecting antibodies against Haemophilus,
 CC HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
 CC non-typeable strains of Haemophilus via hybridisation reactions. The
 CC present sequence represents an HMWA protein from a non-typeable strain of
 CC H. Influenzae.

SQ Sequence 1228 AA;

Query Match 5.6%; Score 106; DB 21; Length 1228;
 Best Local Similarity 20.2%; Pred. NO. 1.3;
 Matches 91; Conservative 67; Mismatches 165; Indels 128; Gaps 21;

QY 11 FFSVLFETFAVSAD-----KIPGDESITNIFGPRDRNESSPRKHN-----LNHITAY 58
 DB 271 fndvtvlnaasavvfnkplvdkvln-----gnhllkgnlsvlvggdvnhfnas 324
 QY 59 SESHTLYDRMTFLCLSHN-----TLNGA 82
 DB 325 ssnlytly-----vllsegnfaasgsslkfkfseghaafctkndllnatgnslngv 380
 QY 83 CPTSENPSSSSYSGET-----NITLOPTEKRSILKRELQIKGYKOLLFKSYNC----- 130
 DB 381 agldsnlksllanfkfegntllaadkpklelkgnlvtegegnvclrsangndksa 440
 QY 131 -----PSGLTNSAHFNCNKN-AASGASLYLYPAGELKNPF--GGIWDATLKL 178
 DB 441 lsiqnvtnkgnlvtcgsaineknlvtegsakflamp-----nysfivsglfdqgksn 495
 QY 179 VK-----RRYSCTGYTINIIKLT-----KGNIOIWLQFQSDARVDNLPTGAG 227
 DB 496 lslakgaifkldientglnltkdsnhltikgnl-----tnkgdlnicnngdn 547
 QY 228 T--YIGRNSVDMCFYDGYSTNSS-----LEIRFQDNPKSD-GKFFLRKINDTKEI 277

DB 548 teiqign-----lsqkegnltsdskvnterltlkagvngdsneatsanltktel 604
 QY 278 AVTSLLAG---KSLPTNGTSLNIADAASL-ENNMRTIAVMPETISVVICMPGRLO 333
 DB 605 klndlnisgfnkaeltakdsnltlignsdagndakkvlfsvnkdkiskas--dhnvt 662
 QY 334 LDAKVE---NPEAGQYMGNIWVTFPSSQTL 361
 DB 663 lnskvetsgdcdstdedggnmhtgltaknv 693

RESULT 15
 AAR41731
 ID AAR41731 standard; Protein; 1338 AA.

AC AAR41731;
 DT 26-APR-1994 (first entry)

DE High molecular weight protein 3 (HMW3).

KW HMW; high molecular weight protein; virus; vaccine; influenza;
 KW epitope; immunity; haemophilus influenzae.

OS Haemophilus influenzae.

PN W09319090-A.

XX 30-SEP-1993.

PD 16-MAR-1993; 93WO-US02166.

PF 16-MAR-1992; 92GB-0005704.

XX 16-MAR-1992; 92GB-0005704.

PA (BARE/) BARENKAMP S J.

PI (INRM) INSERM INST NAT SANTE & RECH MEDICALE.

DR WPI: 1993-320683/40.

XX N-PSDB; AAQ49510.

PT High molecular weight surface proteins - of non-typeable

PT haemophilus which exhibit immunogenic properties

CC Claim 5; Figure 10; 100pp; English.

CC The isolation and purification of the high molecular weight protein
 CC enables the identification of the major protective epitopes of the
 CC protein by conventional epitope mapping. These epitopes can then be
 CC synthesised using standard techniques and incorporated into fully
 CC synthetic or recombinant vaccines.

SQ Sequence 1338 AA;

Query Match 5.4%; Score 101; DB 14; Length 1338;
 Best Local Similarity 21.5%; Pred. NO. 4.3;
 Matches 90; Conservative 52; Mismatches 172; Indels 104; Gaps 18;

QY 21 VSADKIPGDESITNIFGPRDRNESSPRKHN-LNHHI-TAYSESHTLYDRMTFLCL----- 74
 DB 817 vtfdkv-----kdkistidglnvlnsevktsngsnagnstgltsakdv 864
 QY 75 -----SHNTLN-----GACPTSENPSSSSYSGETNITLO-----FTEKR 108
 DB 865 tvnnvntshklnltnsaagvntkgtltnatgsvetagnlkgntlqnvtylate 924
 QY 109 SLIKRELOI---KGYKOLLFKSYNCPGLTNSAHFNCNKNNAASGASL----- 153
 DB 925 nlvttenavlnatstgnlstkgtlkglesstgnvni---tasgnllkvsnlgtgvt 981

```

Oy 154 -----XLYIPAGELKMLPRGGIMADTTLKLRVKRRSRSEVGYTNTITKLIDGNIOW 207
Db 982 vtaagaalttgaagtsiatatgnaaitcttggdingkxwessgsvllvatgaalavniisgn 104111
Oy 208 LPOKSDARVADLMLPTGGYIGRNSVDMCFDYGTJNSSLSLEIRFDONN---PKSDGK 264
Db 1042 tvltladsq---klstsvgtatngtmsvlt-----saqsgllegltisgnlvnvtlaagtd 10922
Oy 265 FYL-RKINDOTKELAYTLS-----LLAGKSLITPNGISLMIADAASLEIYMNW--RTTA 315
Db 1093 lltinasvkaakngaaltlttaesgklcttgcgssltssngtlltktadsslaqnihaanvlt 11522
Oy 316 VTMPEISV----PYLCPMGRLOL---DAVENPEAG-----QYMNINVTMPSS 358
Db 1153 ntgttlttgdsrknaatsgcltlnakdakilaaasddrtvnaatnaasgntaktss 1210

```

Search completed: July 1, 2002, 11:24:10
Job time: 317 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 1, 2002, 11:19:28 ; Search time 13.09 Seconds

(without alignments)
673.616 Million cell updates/sec

Title: US-09-839-894-10

Perfect score: 1866
Sequence: 1 MNKILFIFLTFSSVLTFFFA.....EAGQYMGNIWTFPSQTL 361Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 200000000Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents_AA*
1: /cgn2_6/prodata/2/1aa/5A_COMB.pep.*
2: /cgn2_6/prodata/2/1aa/5B_COMB.pep.*
3: /cgn2_6/prodata/2/1aa/6A_COMB.pep.*
4: /cgn2_6/prodata/2/1aa/6B_COMB.pep.*
5: /cgn2_6/prodata/2/1aa/PCTUS_COMB.pep.*
6: /cgn2_6/prodata/2/1aa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1536.5	81.5	360	2	US-08-483-101-17 Sequence 17, Appl
2	988.5	52.4	363	2	US-08-483-101-16 Sequence 16, Appl
3	927.5	49.2	364	2	US-08-483-101-5 Sequence 5, Appl
4	116.5	6.2	1600	2	US-08-617-697-10 Sequence 10, Appl
5	114	6.0	650	3	US-08-362-525-2 Sequence 2, Appl
6	111.5	5.9	1529	2	US-08-728-470-10 Sequence 10, Appl
7	111.5	5.9	1529	4	US-08-719-641-10 Sequence 10, Appl
8	106	5.6	671	2	US-08-737-716-13 Sequence 13, Appl
9	101	5.4	1338	2	US-08-728-470-9 Sequence 9, Appl
10	101	5.4	1338	4	US-08-719-641-9 Sequence 9, Appl
11	101	5.4	1599	2	US-08-617-697-9 Sequence 9, Appl
12	99	5.2	1848	4	US-08-296-791-6 Sequence 6, Appl
13	99	5.2	1848	5	PCT-US95-10661A-6 Sequence 6, Appl
14	94.5	5.0	1912	3	US-08-409-995-4 Sequence 4, Appl
15	94.5	5.0	1912	3	US-08-685-467-4 Sequence 4, Appl
16	94.5	5.0	2353	4	US-09-377-155-33 Sequence 33, Appl
17	94.5	5.0	2353	4	US-08-913-942-4 Sequence 4, Appl
18	94.5	5.0	2353	4	US-09-669-974-33 Sequence 33, Appl
19	94.5	5.0	2354	4	US-09-268-347-47 Sequence 47, Appl
20	94.5	5.0	2411	4	US-09-268-347-36 Sequence 36, Appl
21	93	4.9	1167	1	US-08-485-568A-6 Sequence 6, Appl
22	93	4.9	1167	2	US-08-590-554A-6 Sequence 6, Appl
23	93	4.9	1167	2	US-09-184-223-6 Sequence 6, Appl
24	91.5	4.9	869	2	US-08-483-101-15 Sequence 15, Appl
25	91	4.8	1168	1	US-08-620-717A-9 Sequence 9, Appl
26	90.5	4.8	1328	3	US-08-781-891-76 Sequence 76, Appl
27	89	4.7	465	3	US-08-948-997-6 Sequence 6, Appl

28	89	4.7	465	4	US-09-348-817A-6 Sequence 6, Appl
29	88	4.7	1394	4	US-08-296-791-2 Sequence 2, Appl
30	88	4.7	1394	5	PCT-US95-10661A-2 Sequence 2, Appl
31	87.5	4.6	353	3	US-08-650-172-13 Sequence 13, Appl
32	87.5	4.6	553	4	US-09-375-419-13 Sequence 13, Appl
33	87.5	4.6	789	1	US-08-431-080-20 Sequence 20, Appl
34	87.5	4.6	789	1	US-08-938-534-20 Sequence 20, Appl
35	86.5	4.6	1003	1	US-07-991-867B-6 Sequence 6, Appl
36	86.5	4.6	1003	1	US-08-107-755A-6 Sequence 6, Appl
37	86.5	4.6	1003	2	US-08-544-332-6 Sequence 6, Appl
38	86.5	4.6	1536	1	US-08-038-682-2 Sequence 2, Appl
39	86.5	4.6	1536	1	US-08-302-832-2 Sequence 2, Appl
40	86.5	4.6	1536	2	US-08-530-198-2 Sequence 2, Appl
41	86.5	4.6	1536	2	US-08-469-880-2 Sequence 2, Appl
42	86.5	4.6	1536	2	US-08-728-470-2 Sequence 2, Appl
43	86.5	4.6	1536	2	US-08-617-697-2 Sequence 2, Appl
44	86.5	4.6	1536	4	US-08-719-641-2 Sequence 2, Appl
45	86	4.6	873	2	US-08-912-129A-61 Sequence 61, Appl

ALIGNMENTS

RESULT 1
US-08-483-101-17
Sequence 17, Application US/08483101
Patent No. 5932715
GENERAL INFORMATION:
APPLICANT: Scott, June R.
APPLICANT: Froehlich, Barbara
TITLE OF INVENTION: CS2 Proteins and Coding Sequences
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Greenlee and Winner, P.C.
STREET: 5370 Manhattan Circle, Suite 201
CITY: Boulder
STATE: Colorado
COUNTRY: US
ZIP: 80303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/483,101
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Ferber, Donna M.
REGISTRATION NUMBER: 33878
REFERENCE/DOCKET NUMBER: 6-95
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 489-8080
TELEFAX: (303) 499-8089
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 360 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: protein
HYPOTHEICAL: NO
US-08-483-101-17

Query Match 81.5%; Score 1536.5; DB 2; Length 360;
Best Local Similarity 80.1%; Pred. No. 5.5e-150;
Matches 289; Conservative 33; Mismatches 38; Indels 1; Gaps 1;
OY 1 MNKILFIFLTFSSVLTFFAVSADKIPGDSITITPCPRDNSSPKHNLNNITAYSE 60
|||||

Db 1 MNKILFIETFFSSGFFTFPAVSADKNPSENMTNTICPHDRGSSPIYINILNSYLAYNG 60
QY 61 SHTLYDMRTFLCLSSHNTLNGACPTSENPSSSVSGETNTITLOFTEKRSILIKRELQIKGY 120
Db 61 SHHLYDMRTFLCLSSQNTLNGACPSDAPGTATIDGETNTITLOFTEKRSILIKRELQIKGY 120
QY 121 KOLLFNSVNPSSQTLNLSAIFNCKNAASGLYLYTPAGELKNLPFGGIMWDTLKIRVK 180
Db 121 KOLFLFNANCPSKLALNSHFCOCNREASCATISLYTPAGELKNLPFGGVWVNLKLNK 180
QY 181 RRTSETGTYTNTITIKLTKGNIOIWLPOFKSDARYDLNLRPFGGTYIGRNSVDMCFY 240
Db 181 RRTDTTYGTTITNTITVLTGKNIQIWLPOFKSNARYDLNLRPFGGTYIGRNSVDMCFY 240
QY 241 DGXSTNSSLEIRFODNPKSDGKFLRKINDTKETIAYTLSTLLAGKSLPTNGTSLNI 300
Db 241 DGXSTNSSLEIRFODNPKSDGKFLRKINDTSKELVYTLSTLLAGKKNLTPNGQALNI 300
QY 301 ADAASLETNNRITAYTMPEISVPLCWPGRLQDLAKVNPENAGQYMGINVTFTPSQOT 360
Db 301 -NTASLETNNRITAYTMPEISVPLCWPGRLQDLAKVKNPEAGEYSGIINVTFTPSSSS 359
QY 361 L 361
Db 360 L 360
RESULT 2
US-08-483-101-16
; Sequence 16, Application US/08483101
; Patent No. 5932715
; GENERAL INFORMATION:
; APPLICANT: Scott, June R.
; APPLICANT: Froehlich, Barbara
; TITLE OF INVENTION: CS2 Proteins and Coding Sequences
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee and Winner, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: US
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,101
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33878
; REFERENCE/DOCKET NUMBER: 6-95
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 499-8080
; TELEFAX: (303) 499-8089
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 363 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
US-08-483-101-16

Query Match 52.4%; Score 988.5; DB 2; Length 363;
Best Local Similarity 55.0%; Pred. No. 1.7e-93;

Matches 204; Conservative 53; Mismatches 93; Indels 21; Gaps 11;
QY 3 KILFIETFFSSVLETFPAVSADKIPGDE--SITNIF-GPR-DNNESSPKHININITAY 58
Db 2 KRIIFIL-----SIFSAVVSAGRYPEYTVGNLTKSQAPRLDSSVSPITNIFTNVHAG 57
QY 59 SESHLYDMRTFLCLSSHNTLNGACPTSENPSSSVS-GETNTITLOFTEKRSILIKRELQI 117
Db 58 SLHSILYDRIVFLCTSSSNVNGACPTI--GTSVQYGTITITLOFTEKRSILIKRNINIL 114
QY 118 KGYKOLLFNSVNPSSG--LTILNSAHFNCKNA-ASGLYLYTPAGELKNLPFGGIMWDT 174
Db 115 AGMKKPLTWEMQSCDFSNLWLNKSMSCAGHANGTLLNLYTPAGELKNLPFGGIMWDT 174
QY 175 LKLRVRRYSET----YGTITNTITIKLTDKGNIOIWLPOFKSDARYDLNLRPFGGTYI 230
Db 175 LIIRLS-RGEVSVSTHGNTVITVDLIDKGNIOIWLPOFHSNRPVLDLNRIGYKKS 233
QY 231 GRNSVDMCFYDGXSTNSSLEIRFODNPKSDGKFLRKINDTKETIAYTLSTLLAGKSL 290
Db 234 GSNSLDMCFYDGXSTNSDSNVIKFODDNPNSSEVNLKIG-CTEKLPYAVSLIMGEKIF 292
QY 291 TPTNGTSLNIDAAASLETNNRITAYTMPEISVPLCWPGRLQDLAKVNPENAGQYMGIN 350
Db 293 YPVNGQSFITINDSSVLETNNRITAYTMPEVNPVPLCWPARLLNADVNNPEAGQYMGIN 352
QY 351 NVTFTPSQOTL 361
Db 353 KITFTPSQOTL 363
RESULT 3
US-08-483-101-5
; Sequence 5, Application US/08483101
; Patent No. 5932715
; GENERAL INFORMATION:
; APPLICANT: Scott, June R.
; APPLICANT: Froehlich, Barbara
; TITLE OF INVENTION: CS2 Proteins and Coding Sequences
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee and Winner, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: US
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,101
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33878
; REFERENCE/DOCKET NUMBER: 6-95
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 499-8080
; TELEFAX: (303) 499-8089
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 364 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-483-101-5

```

Query Match          49.2%; Score 927.5; DB 2; Length 364;
Best Local Similarity 48.1%; Pred. No. 3,3e-87;
Matches 176; Conservative 65; Mismatches 118; Indels 7; Gaps 3;

OY 1 MNKLEFLEFFSSVLETFVAVSADKIPGDESTINIEGPRDRN--ESSPKHNINNHITAY 58
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1 LKVIYFVLSMFLCSQYVQGSMTNVEAGSINKTESIGPIDRSAASVPAHYIFHEHAGY 60

OY 59 SESITLYDRKTFELCSLHNTLNGACPTSENPSSSVSGENITLOFTEKSLKRELQIK 118
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 61 NKHSLSFDRMTFLCMSSTDAKSGCAPGENSKSS--OGETNIKILFPEKSLAKRTLNK 118

OY 119 GYKOLFEKSVNC--PGSLNLSAHFNCKNAASGASLYIYPAGELKNLPFGIMPATL 175
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 119 GYKRFLESORCHYVDKMLNSHTKVCVSGFTRGVDFITYIPOGELDGLITGIMPATL 178

OY 176 KLRVRRYSEYGYTYTITIKLTKGNQIWLPOFKSARVDLNLAPTGSGGYIGNSV 235
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 179 ELRVKRRHYDNGHYKYNITVDLTKGNIQVMPKPHSDRIDNLNRPENKSGSNVL 238

OY 236 DMCFTYDGYSTNSSLERFODNNPKSDGKFFYLKINDTKEIATYLSLLAGSLTPFG 295
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 239 EMLCYDGYSHSOSIEKRFDDSGTGNENYNLKGTGELKLPYKSLILGREFEYPNG 298

OY 296 TSLNIADAASLETNMRITAVTMEISVPLCWPGRLOLDAKVENPEAGQVGNINVTFT 355
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 299 EAFINDTSSLFINMNRKIKSVSLPOLISIPVLCWPANLTFMSELNPEAGEISGLINVTFT 358

OY 356 PSSQTL 361
   : : : : :
Db 359 PSSSSL 364

RESULT 4
US-08-617-697-10
; Sequence 10, Application US/08617697
; Patent No. 5977336
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5977336-Typeable Haemophilus
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Matzare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/617, 697
; FILING DATE: 01-APR-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 05-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-557
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 10:

```

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 1600 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-617-697-10

Query Match          6.2%; Score 116.5; DB 2; Length 1600;
Best Local Similarity 21.6%; Pred. No. 0.014;
Matches 74; Conservative 60; Mismatches 130; Indels 79; Gaps 19;

OY 48 HNLNHNITAVSSHTLYDMMTFELCSLHNTLNGACPTSENPSSS--SVSGETITL 102
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 816 YNEYSKH--AINSSHNL-----TILGNAVTLGG-----ENSSSITGNTINITKNAVTL 862

OY 103 QF-----TEKSLIKRELQIKYKQLFEKSYNCPGSLTL-----NSAHFNCN-- 144
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 863 QADTSNNTGLKRRKTLTLGNISVEGNLSLIGANANIVGNLSIAMEDSTFKGEASDNLNITG 922

OY 145 ---KNAASGASLYIYPAGELKNLPFGIMPATLKLVRKRYSETYGT--YTINITIKLTD 200
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 923 TFTNNGTANINIKGVYALGDIINKK--GG-----LNTTNASGTOKTIIINGNT---NE 970

OY 201 KGNQIQLPQFKSPARDL--NLAPTGSGGYIGNSVDMCFYDGYSTNSSLERQ--- 255
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 971 KGDINT--KNIKDAEALQIGNTSQKEGNTLTISDKVNT-----TNQITTAGVGGGR 1021

OY 256 -DNNPKSDGKFFYLKINDTKEIATYLSLLAG---KSLPTNGTSLNIADAASLETNWN 311
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1022 SDSSEANANLTLT-----TKELKLAGDLNLSGFENKAEITAKNGSDTLTINASGNNDAK 1076

OY 312 RITAVTMEISVPLCWPGR--RLQOLDAKVENPEAGQVGNINVT 353
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1077 K--VFEDKVKDSKISTGDHNVTLNLSKVTSNGSSNAGNDNST 1116

RESULT 5
US-08-362-525-2
; Sequence 2, Application US/08362525
; Patent No. 6027910
; GENERAL INFORMATION:
; APPLICANT: KILS, FRANCISCUS M.
; APPLICANT: SCHREIDER, MARTEN P.
; APPLICANT: TOSCHKA, HOLSER Y.
; APPLICANT: VERRIPS, CORNELIS T.
; TITLE OF INVENTION: PROCESS FOR IMMOBILIZING ENZYMES TO THE
; TITLE OF INVENTION: CELL WALL OF A MICROBIAL CELL BY PRODUCING A FUSION
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.
; STREET: 1100 New York Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3918
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/362,525
; FILING DATE: 04-JAN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 92202080.5
; FILING DATE: 08-JUL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 92203899.7
; FILING DATE: 14-DEC-1992
; PRIOR APPLICATION DATA:

```

APPLICATION NUMBER: PCT/EP93/01763
FILING DATE: 07-JUL-1993
ATTORNEY/AGENT INFORMATION:
NAME: KOKULIS, PAUL N.
REGISTRATION NUMBER: 16,773
REFERENCE/DOCKET NUMBER: 213289/T7020(V)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 861-3000
TELEFAX: (202) 822-0944
TELEX: 6714627 CUSH
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 650 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-362-525-2

Query Match 6.0%; Score 114; DB 3; Length 650;
Best local Similarity 20.9%; Pred. No. 0.0057;
Matches 98; Conservative 52; Mismatches 147; Indels 172; Gaps 23;

QY 4 ILPIFLFESSVL-----FTFA-----VSADKIPG-----DESTINIGPPDRNE--- 43
DB 8 ILMFLSLALASAININDIFPSNLEIPLTANKOPDQGWATPFPSIADASSIREGDEFTL 67
QY 44 SSP---KHNILNHNHTA-----YSESHLYDRMTFLC-----LSNHTLN 80
DB 68 SMHVYRIKILNSOTATITSLADGTEFAKCYVSOQAAYILENTTFTCTAONDISVTID 127
QY 81 GACPTSENPSSSVSGETNITTOFTTEKRLIKRELOIKGYKOLLFRKSVNC-PSGLTLNSA 139
DB 128 GSITFSLFNSDGSSSYEYL-----ENAKFKSGPMLVKGNQMSDVVNPDPAFTENVF 182
QY 140 HFNCNNKASGASLYL--YIPAGELKNLPPGGIWDATLKLKVRKRRISSETYGTITNITIK 197
DB 183 HSGRSRIGYGSFESYHLGMCPCPNFY---LGG---TEKI-----DVSNNVNDL- 225
QY 198 LTFKGNQI-----MLPQKSDARVDLNLPRGGGTYIGRN---SYDMCFYDG--- 242
DB 226 --DCSSVOYVSSNDFNDMFPPOSTINDTADY-----TCFSSNMTTLDEKLYDEML 275
QY 243 -----YSTNSSLE--IRPDNN--PKSDGKF 265
DB 276 WYVALGSLPANVNTIDHALEFOYTCLDTIANTTYATQFSTRREIYQGNLGTASAKSS 335
QY 266 YLRKINDTKET---AYTL-----LILAKSLPTPTNGSLNADAA 304
DB 336 FISTTTTDLTISMTSAYSTGSIIVENGNTTSEVISHVVTSTKLSPTATSLTIAOTS 395
QY 305 SLETPNN-----RTATYMPETISVPVLCWQRL 332
DB 396 IYSTDSNITVGTIHTTSEYISDVEYISRETASTVAAPTSTTGWTCAM 444

RESULT 6
US-08-728-470-10
Sequence 10, Application US/08728470
Patent No. 5928651
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Matzare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/728,470
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-633
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 413-0810
TELEFAX: (703) 413-0810
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1529 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-728-470-10

Query Match 5.9%; Score 111.5; DB 2; Length 1529;
Best local Similarity 20.7%; Pred. No. 0.041;
Matches 72; Conservative 62; Mismatches 127; Indels 87; Gaps 18;

QY 48 HNILNHNHTAYSESHLYDRMTFLCLSHNHTLNGACPTSENPSS-----SVSGETNIRL 102
DB 744 YNEYSKH--AINSSHNL-----TILGNAVTLG-----ENSSSTTGNNITNKNVTL 790
QY 103 QF-----TEKRLIKRELOIKGYKOLLFRKSVNCPGSLTLNSAHFNCKNKAASGLY 154
DB 791 QADTSNMGKLRKRLTLGIVSEGNLSLGAANIVGNLSI--ARDSTFKGAS----- 843
QY 155 LYIPAGELKNLPPGGIWDATLKLKVRKRRISSETYGTITNITIKL----- 199
DB 844 -----DNLNTTGTFTNGTANINIKOGVYKLOGDINNKGGLNTTNASGTOKTIIING 895
QY 200 ---DKNGIQIWLPPKSDARVDL--NLPRPTGGGTYIGRNSVDMCFYDGYSTNSSLEIR 253
DB 896 NITNEKGDLLT--KNIKADAEIOIGNISOKESNLTISSDKYI-----TROITIKAG 946
QY 254 FQ---DNNPKSDGKFLYLRKINDTKETIAYTSLLAG---KSLPTNGTSLNADAA 306
DB 947 VEGGRSDSSEANANLTIQ---TKELKLAGDLINISFNKAKAITAKNSDILLIGNASGG 1001
QY 307 ETWNNKITAVTMEISVPVLCWPG-RLQDPAKYENPEAQYKMNINVT 353
DB 1002 NADAKR---VTFDKVADSKISTDGHNVTLNSEVKTSGSSMAGDNST 1046

RESULT 7
US-08-719-641-10
Sequence 10, Application US/08719641
Patent No. 6218141
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 6218141-Typeable Haemophilus
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:

ADDRESSEE: Shoemaker and Mattare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/719,641
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 16-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Belkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-625
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1529 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-719-641-10

Query Match 5.9%; Score 111.5; DB 4; Length 1529;
Best Local Similarity 20.7%; Pred. No. 0.041;
Matches 72; Conservative 62; Mismatches 127; Indels 87; Gaps 16;
QY 48 HNLNHNITAVSESHLYDRMTFLCLSSHNTLNGACPTSENPSSS-----SVSGFTNITL 102
DB 744 YNEYSKH--AINSSHNL-----TILGDNVTLG-----ENSSSITGNTNITNKANVTL 790
QY 103 QF-----TEKRLIRRELQIKYKQLLEKSVNCPGSLTINSAHFNCNKAAGASLY 154
DB 791 QADTNSNMTGKRLTLGNISVEGNLSLTGANANIVGNLSI--AEDSTFGKAS----- 843
QY 155 LYIPAGELKNLPFGGIW--DATLKLVRKRYSEYGTYT-----INITKLT----- 199
DB 844 -----DNLNTGTFTNNGTANINIKOGVYKLOGDINNKGGLTITTNASTGTITING 895
QY 200 ----DKGNIQIWLPOFKSDARVDL--NLPTGGGTYIGRNSVDMCFYDGYSTNSSLEIR 253
DB 896 NITNEKGLNLT--KNIKADAELQIGCNISQKEGNLTSSDKVNI-----TNOQTTIRAG 946
QY 254 FQ-----DNMPSDGGFYLRKINDPKETAYTILSLLAG---KSLTPNIGTSLINTADA 306
DB 947 VEGGRSDSESEANLTLQ-----TKELKLAGDLNISGFNKAETIRAKKNGSDDLTI 1001
QY 307 ETNNMIRITAVMPEISVPLCMPG--RLQDLAKVENPEAGGYWGNINVT 353
DB 1002 NADAK-----VTFDKYKDSKISTDGHNTVLNSEVKTSGNSSNAGNDNST 1046

RESULT 8
US-08-737-716-13
; Sequence 13, Application US/08737716

Patent No. 5955258
GENERAL INFORMATION:
APPLICANT: Giorde BUIST
APPLICANT: Gerard VENEMA
APPLICANT: Jan KOK
TITLE OF INVENTION: Process for the lysis of a culture of lactic
TITLE OF INVENTION: acid bacteria by means of a lysis, and uses of the resultin
TITLE OF INVENTION: lysed culture.
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pillsbury Madison & Sutro, L.L.P.
STREET: 1100 New York Avenue, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005-3918
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Microsoft Word
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/737,716
FILING DATE: 22-APR-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/NL95/00170
FILING DATE: 12-MAY-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 94201353.3
FILING DATE: 12-MAY-1994
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 671 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Streptococcus faecalis
IMMEDIATE SOURCE:
CLONE: Fig.5a (S. faecalis)
US-08-737-716-13

Query Match 5.6%; Score 106; DB 2; Length 671;
Best Local Similarity 23.6%; Pred. No. 0.04;
Matches 73; Conservative 39; Mismatches 103; Indels 94; Gaps 17;
QY 41 RNESPKNHT--LNHNITAVSESHLYDRMTFLCLSSHNTLNGACPTSENPSSSV----- 94
DB 309 RYATDPSPYNAKLNINVTAY--NLTOYDTPSSGNTGGTVPNGTGSNNQSTNYYTVK 366
QY 95 SGET--NITLQF-----TEKRL--IKRELQIKYKQLLEKSVNCPGSLTINSAHFNCNKN 146
DB 367 SODTLNKIAAQAQGYVAVANLRKSWNGISGDLIFYGQRLVKKGA---SGNTGSGNGSGNNN 423
QY 147 AASGASLYLYIPAGELKNLPFGGIWDAFLKLVRKRYSEYGTYTINITIKLTDKGNIOI 206
DB 424 -QSGTNTFYTVKSGDTLN-----KIAAQYG-----VTY-----ANLRS 455
QY 207 WLPOFKSDARVDLNL-----RPTGGGTYIGRNSVDMCFYDGYSTNSSLEIRFQNN 258
DB 456 W-----NCSGDLIPVQOKLYVKKTSNT-----GSSNGS-----NNN 491
QY 259 PKSDGKFLRKINDPKETEA--YTLS-----LLAGSLTPNIGTSLINTADA 303
DB 492 QSGTNTFYTVKSGDTLNKIAAQAQGYVAVANLRKSWNGISGDLIFAGOKITVKKGTSGNTGGS 551
QY 304 ASLETNMR 312
DB 552 SNGSGNNNO 560

```

: RESULT 9
: US-08-728-470-9
: Sequence 9, Application US/08728470
: Patent No. 5928651
: GENERAL INFORMATION:
: APPLICANT: Barenkamp, Stephen J
: TITLE OF INVENTION: High Molecular Weight Surface Proteins
: TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus
: NUMBER OF SEQUENCES: 10
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Shoemaker and Mattare, Ltd.
: STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
: STREET: Bldg. 1
: CITY: Arlington
: STATE: Virginia
: COUNTRY: U.S.A.
: ZIP: 22202-0286
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/728,470
: FILING DATE:
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/302,832
: FILING DATE: 16-MAR-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US PCT/US93/02166
: FILING DATE: 16-MAR-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: GB 9205704.1
: FILING DATE: 16-MAR-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Berkstresser, Jerry W
: REGISTRATION NUMBER: 22,651
: REFERENCE/DOCKET NUMBER: 1038-633
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703) 415-0810
: TELEFAX: (703) 415-0813
: INFORMATION FOR SEQ ID NO: 9:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1338 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: US-08-728-470-9

```

Query Match	5.4%;	Score 101;	DB 2;	Length 1338;
Best Local Similarity	21.5%;	Pred. NO. 0.4;		
Matches	90;	Conservative	52;	Mismatches 172; Indels 104; Gaps 18;
Oy	21	VSADKIPGDESIITNIEGPRDNESSPKHNI-LNNHI-LAYSESHLYDREMLCLS----	74	
		! : ! :	! : ! :	! : ! :
Db	817	VTFEDV-----KDSKISDGHVLTNTSEYKTSNGSSNAGNDNSTGLTISAKDY	864	
Oy	75	-----SHNTLN-----GACPTSENPPSSSVSGENITLO-----FTEKR	108	
		! ! ! !	! ! ! !	! ! ! !
Db	865	TYNNNVNTHSKITINISAAAGNVTTKEGTTINATGGSVEYTAOMGTIKGNITTSQNVITATE	924	
Oy	109	SLIKRELQI---KGYKOLLKFSVNCPSGLTINSAHFNCNKMAAGAST-----	153	
		: ! : !	: ! : !	: ! : !
Db	935	MLVTEMAVINATSGEIVNISTKGTGCIKGIESTSGNVNI---TASGNITLKVANITGQDVT	981	
Oy	154	-----LYIPAPGELKNLPFGCIMPATLKLRYKRRRSEYGYTYNTITKLDKGIQIOW	207	
		! ! ! !	! ! ! !	! ! ! !
Db	982	VTADAGALTTTAGTITISATNGTANNTTKRGDINGKEVSESSGVTLVATGATILAVGNISGN	1041	
Oy	208	LPFKSDAEVDLNLPRPTGGGITYIGKNSVDMCFYDGTISTINSSSLEIRFDNN---PKSDGK	264	

```

Db      1042  TTTTADSG--KLTSTVSGTNGTNSVTT-----SQSGDIEGHSIGNTVNTATSPGD 10922
QY      265  FYL-RKINDOTKEAYTL-----LLAGKSITPNGISLNIADAASLEIWN--RTA 315
Db      1093  LTTINSARVAKNGAALITAEESKLTPTQSGSITSSNGOTTLTADSSIAGINANAYTL 1152
QY      316  VTMEISV---PYLCPGRLOT---DAVEVNEAG-----QYMGINITFPSS 358
Db      1153  NTGTGTLTTTGDSKINATSGTTLTNADAKIDCAASDRVNVATNTAASGANTYATSS 1210

```

```

1 RESULT 10
2 US-08-719-641-9
3 : Sequence 9, Application US/08719641
4 : Patent No. 6218141
5 :
6 : GENERAL INFORMATION:
7 : APPLICANT: Barenkamp, Stephen J
8 : TITLE OF INVENTION: High Molecular Weight Surface Proteins
9 : TITLE OF INVENTION: of No. 6218141-Typeable Haemophilus
10 : NUMBER OF SEQUENCES: 10
11 :
12 : CORRESPONDENCE ADDRESS:
13 : ADDRESS: Shoemaker and Mattare, Ltd.
14 : STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
15 : STREET: Bldg. 1
16 : CITY: Arlington
17 : STATE: Virginia
18 : COUNTRY: U.S.A.
19 : ZIP: 22202-0286
20 :
21 : COMPUTER READABLE FORM:
22 : MEDIUM TYPE: Floppy disk
23 :
24 : COMPUTER: IBM PC compatible
25 : OPERATING SYSTEM: PC-DOS/MS-DOS
26 : SOFTWARE: Patentln Release #1.0, Version #1.30
27 : CURRENT APPLICATION DATA:
28 : APPLICATION NUMBER: US/08/719,641
29 : FILING DATE:
30 :
31 : CLASSIFICATION: 530
32 :
33 : PRIOR APPLICATION DATA:
34 : APPLICATION NUMBER: US 08/302,832
35 : FILING DATE: 16-SEP-1994
36 :
37 : PRIOR APPLICATION DATA:
38 : APPLICATION NUMBER: US PCT/US93/02166
39 : FILING DATE: 16-MAR-1993
40 :
41 : PRIOR APPLICATION DATA:
42 : APPLICATION NUMBER: GB 9205704.1
43 : FILING DATE: 16-MAR-1992
44 :
45 : ATTORNEY/AGENT INFORMATION:
46 : NAME: Belkstresser, Jerry W
47 : REGISTRATION NUMBER: 22,651
48 : REFERENCE/DOCKET NUMBER: 1038-625
49 : TELECOMMUNICATION INFORMATION:
50 : TELEPHONE: (703) 415-0810
51 : TELEFAX: (703) 415-0813
52 :
53 : INFORMATION FOR SEQ ID NO: 9:
54 : SEQUENCE CHARACTERISTICS:
55 : LENGTH: 1338 amino acids
56 : TYPE: amino acid
57 : STRANDEDNESS: single
58 : TOPOLOGY: linear
59 :
60 : US-08-719-641-9

```

```

Query Match          5 4%: Score 101: DB 4: Length 1338;
Best Local Similarity 21.5%: Pred. NO. 0.4;
Matches 90; Conservative 152; Mismatches 172; Indels 104; Gaps 18;

QY      21 VSADKIPGDESITNIFGRDRNESSPKHNI-LNNHI-TAYSSEHTLYDRMFELCS---- 74
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      817 VTPEKV-----KDSKISITGDHNVTLNSEVTKISNGSSNAGNDSTGLTISAKDV 864
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      75 -----SHNTLN-----GACPTISNPSSSSVSGETNITLQ-----FTEKR 108
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Db 865 TVNNVNTSHKNTINISAAGNVTTKEGTTINATGSEVTAQNGTIKGNITSQNVYATATE 924
QY 109 SLIKREIQI-----KGYKQLFKSVNCPGSLTNSAHFNCKNKAASGL----- 153
Db 925 NLVTTENAIVNATSGVYVNSTKGTGDKIGIESTSGVNI---TASGNTLKVNSITGDDVT 981
QY 154 -----YLYIPAGELKNLPFGGIMDATLKLVRKRRSEYGYTYTINTIKLTDKNGIQIW 207
Db 982 VTADAGALTTTASTISATIGMANITTKTDINGKEVSSGSLVATGATLAVAGNSGN 1041
QY 208 LPOFKSDARVDLNLPRPGGTYIGRNSVDMCFYDGYSTNSSLEIRQDNN---PKSDGK 264
Db 1042 TVTITADSG---KLTSTVGSTINGTNSVT-----SSQSGDIEGTISGNTVAVTASTGD 1092
QY 265 FYL-RKINDTKETIAYTL-----LLAGKSLPTNGTSLNIAAASLETNNN--RITA 315
Db 1093 LITGNSAKVEAKNGAATLTPESGKLTQTQSSITSSNGQTTTLAKOSSISAGINIAANVT 1152
QY 316 VTMPETISV---PYLCMPGRLOL---DAKVENPAG-----QYGNINVTFTPS 358
Db 1153 NTGTGTLTTGDSKINATSGTLTINAKDAKLDGAASGDRIVVATNATNAGSGNVATKTS 1210

RESULT 11
US-08-617-697-9
; Sequence 9, Application US/08617697
; Patent No. 5977336
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5977336-Typeable Haemophilus
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Maltare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/617,697
; FILING DATE: 01-Apr-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 05-Oct-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-Mar-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-557
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1599 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-617-697-9

Query Match 5.4%; Score 101; DB 2; Length 1599;
Best Local Similarity 21.5%; Pred. No. 0.54;

Matches 90; Conservative 52; Mismatches 172; Indels 104; Gaps 18;
QY 21 VSADKIPGDESTIRNIEPRORNESSPKHNI-LNNHI-TAISSEHTLYDRRTFLCS---- 74
Db 1077 VTEDKV-----KSKISTDGHNVTLNSEVTSNGSSNAGDNSTGLTISAKDV 1124
QY 75 -----SHNTLN-----GACPTSENPSSSSVSGETNITLO-----FEKR 108
Db 1125 TVNNVNTSHKNTINISAAGNVTTKEGTTINATGSEVTAQNGTIKGNITSQNVYATATE 1184
QY 109 SLIKREIQI-----KGYKQLFKSVNCPGSLTNSAHFNCKNKAASGL----- 153
Db 1185 NLVTTENAIVNATSGVYVNSTKGTGDKIGIESTSGVNI---TASGNTLKVNSITGDDVT 1241
QY 154 -----YLYIPAGELKNLPFGGIMDATLKLVRKRRSEYGYTYTINTIKLTDKNGIQIW 207
Db 1242 VTADAGALTTTASTISATIGMANITTKTDINGKEVSSGSLVATGATLAVAGNSGN 1301
QY 208 LPOFKSDARVDLNLPRPGGTYIGRNSVDMCFYDGYSTNSSLEIRQDNN---PKSDGK 264
Db 1302 TVTITADSG---KLTSTVGSTINGTNSVT-----SSQSGDIEGTISGNTVAVTASTGD 1352
QY 265 FYL-RKINDTKETIAYTL-----LLAGKSLPTNGTSLNIAAASLETNNN--RITA 315
Db 1353 LITGNSAKVEAKNGAATLTPESGKLTQTQSSITSSNGQTTTLAKOSSISAGINIAANVT 1412
QY 316 VTMPETISV---PYLCMPGRLOL---DAKVENPAG-----QYGNINVTFTPS 358
Db 1413 NTGTGTLTTGDSKINATSGTLTINAKDAKLDGAASGDRIVVATNATNAGSGNVATKTS 1470

RESULT 12
US-08-296-791-6
; Sequence 6, Application US/08296791
; Patent No. 6245337
; GENERAL INFORMATION:
; APPLICANT: St. Gene III, Joseph W.
; APPLICANT: Falkow, Stanley
; TITLE OF INVENTION: Haemophilus Adherence and Penetration
; TITLE OF INVENTION: Protein
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobach, Test, Albrighton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/296,791
; FILING DATE: 25-Aug-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Trecartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1848 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; US-08-296-791-6

Query Match 5.2%; Score 99; DB 4; Length 1848;
Best Local Similarity 23.4%; Pred. No. 1.1;
Matches 76; Conservative 56; Mismatches 127; Indels 66; Gaps 20;

QY 19 FAVSADKIGDGSITNIPFPRDRNESSPKHNILNNITAVSES--HTLDRMTFLCLSS- 75
DB 753 FRAITMNVATGNASLVY-----GRNVA-----NITSN--ITASNNAAQVHIGKTDGTVCVRSD 802
QY 76 -----HNT--LNGACPTSENPSSSVSGEFTNT--LQFTEKSLIKRELQIKGYQOLF 125
DB 803 YGVYVCHNSNLSEKALNSFN--TILRGVNLTEANSTLTKANLFGTIQSIGTQVNL 860
QY 126 KSYNCPSSGILLNSAHFNCK--NAAGSASLYLYIPAGELKNLPFGIWDATLKLVRKRRYS 184
DB 861 KE--NSHMLTGNSS--NVQMLNTNG--HILNMQNDAN-----KV 895
QY 185 EYVGYTITITIKLDKNGIQIWLPOFKSDA--RYDLNLAPTGGRY-----IGR--NSVDM 237
DB 896 TTYNTILTVN--SLSGNSFFYTWDFTNKSNKVYVVKASATGFTLQVADKGTGEPHNEL 952
QY 238 CFYDGYSTSSSLEIRFQDNN--PKSDGKFLYLRINDTKEIAYTSLLAGKSLPTNGT 296
DB 953 TLFDSANATRNNEVLVLANGSVDRGAMKYLKLRNVNG--RYDLYNPEVEKRNQTVDTTNT 1010
QY 297 SLN--TADAASLETNNRITAVTMP 319
DB 1011 TENDIQADAPSAQSNNETIARVETP 1035

RESULT 13

PCT-US95-10661A-6
; Sequence 6, Application PC/TUS9510661A
; GENERAL INFORMATION:
; APPLICANT: Washington University, et al.
; TITLE OF INVENTION: Haemophilus Adherence and Penetration Protein
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/10661A
; FILING DATE: 16-AUG-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/296,791
; FILING DATE: 25-AUG-1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Treacartn, Richard F.
; REGISTRATION NUMBER: 31,801
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1848 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; PCT-US95-10661A-6

Query Match 5.2%; Score 99; DB 5; Length 1848;

Best Local Similarity 23.4%; Pred. No. 1.1;
Matches 76; Conservative 56; Mismatches 127; Indels 66; Gaps 20;

QY 19 FAVSADKIGDGSITNIPFPRDRNESSPKHNILNNITAVSES--HTLDRMTFLCLSS- 75
DB 753 FRAITMNVATGNASLVY-----GRNVA-----NITSN--ITASNNAAQVHIGKTDGTVCVRSD 802
QY 76 -----HNT--LNGACPTSENPSSSVSGEFTNT--LQFTEKSLIKRELQIKGYQOLF 125
DB 803 YGVYVCHNSNLSEKALNSFN--TILRGVNLTEANSTLTKANLFGTIQSIGTQVNL 860
QY 126 KSYNCPSSGILLNSAHFNCK--NAAGSASLYLYIPAGELKNLPFGIWDATLKLVRKRRYS 184
DB 861 KE--NSHMLTGNSS--NVQMLNTNG--HILNMQNDAN-----KV 895
QY 185 EYVGYTITITIKLDKNGIQIWLPOFKSDA--RYDLNLAPTGGRY-----IGR--NSVDM 237
DB 896 TTYNTILTVN--SLSGNSFFYTWDFTNKSNKVYVVKASATGFTLQVADKGTGEPHNEL 952
QY 238 CFYDGYSTSSSLEIRFQDNN--PKSDGKFLYLRINDTKEIAYTSLLAGKSLPTNGT 296
DB 953 TLFDSANATRNNEVLVLANGSVDRGAMKYLKLRNVNG--RYDLYNPEVEKRNQTVDTTNT 1010
QY 297 SLN--TADAASLETNNRITAVTMP 319
DB 1011 TENDIQADAPSAQSNNETIARVETP 1035

RESULT 14

US-08-409-995-4
; Sequence 4, Application US/08409995
; Patent No. 5646259
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen I.
; TITLE OF INVENTION: Haemophilus Adhesion Proteins
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/409,995
; FILING DATE: 24-MAR-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-61053/RFT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1912 amino acids
; TYPE: amino acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; US-08-409-995-4

Query Match 5.0%; Score 94.5; DB 1; Length 1912;
Best Local Similarity 19.7%; Pred. No. 3.4;
Matches 61; Conservative 47; Mismatches 117; Indels 85; Gaps 14;

[illegible]

Db 1093 GLATALNLS 1102

5.08; Score 94.5; DB 3; Length 1912;

This Page Blank (uspto)

1	1558.5	82.6	360	2	D56617	cfaf protein prec
2	949.5	50.3	363	2	S49539	COD protein prec
3	930.5	49.3	364	2	S57937	COD protein prec
4	275.5	14.6	359	2	AE0541	probable fimbria
5	120	6.4	1335	2	G90975	probable factor (
6	120	6.4	2660	2	E85892	probable invasin
7	114	6.0	650	2	S22833	alpha-agglutinin
8	109	5.8	2383	2	D64862	probable membrane
9	108	5.7	524	2	S55097	probable membrane
10	108	5.7	2529	2	B64635	toxin-like outer m
11	107	5.7	304	2	A64904	probable fimbrial
12	107	5.7	304	2	C90892	probable adhesin
13	107	5.7	304	2	F85725	probable adhesin
14	106	5.6	671	2	A38109	autolysin - Enter
15	105.5	5.6	843	2	AC2507	hypothetical prote
16	105	5.6	598	2	AB1236	internalin protei
17	104.5	5.5	1441	2	B86807	hypothetical prote
18	103.5	5.5	661	2	B75622	hypothetical prote
19	103.5	5.5	1269	2	A90267	proteinase relate
20	103.5	5.5	2399	2	H71879	toxin-like outer m
21	102	5.4	692	2	G90284	hypothetical prote
22	101	5.4	1238	2	AH0038	probable exported
23	100	5.3	1752	2	T48965	hypothetical prote
24	99.5	5.3	3627	2	C83339	hypothetical prote
25	99	5.2	802	2	AH1580	phenylalanine-tRNA
26	99	5.2	856	2	B81399	probable periplasm
27	99	5.2	1849	2	C41859	Iga-specific metal
28	98.5	5.2	608	2	H64473	hypothetical prote
29	98.5	5.2	4919	2	T31105	hypothetical prote

30	98	5.2	587	2	AC1516	internal protein
31	98	5.2	758	2	S4786	SPR21 protein - yeast
32	98	5.2	1014	2	S37405	cytotoxic necrotic
33	98	5.2	1157	2	S38160	NBP133 protein - y
34	97.5	5.2	1797	2	F69195	cell surface glycop
35	97	5.1	454	2	T01337	hypothetical prote
36	96.5	5.1	2364	2	I40884	cytotoxin L - Clo
37	96	5.1	992	2	T28421	probable DNA-dir
38	95.5	5.1	713	1	ALBSXR	cycloalmodextrin
39	95.5	5.1	895	2	T02597	Mblator-like trans
40	95	5.0	413	1	S28602	translation releas
41	95	5.0	561	2	A84113	transposase (12) E
42	95	5.0	2334	2	S32920	cell wall-associat
43	95	5.0	5188	2	B85547	probable RTX fami
44	94.5	5.0	835	2	E71691	outer membrane as
45	94.5	5.0	858	2	T47223	replication licens

ALIGNMENTS

```

RESULT      1
D56617
cfaf protein precursor - Escherichia coli plasmid NTP113
C:Species: Escherichia coli
C:Date: 05-Jan-1996 #sequence_revision 05-Jan-1996 #text_change 10-Dec-1999
C:Accession: D56617
C:Jordi, B.J.; Millshaw, G.A.; van der Zeijst, B.A.; Gaastra, W.
DNA Seq. 2, 257-263, 1992
A:Title: The complete nucleotide sequence of region 1 of the cfa/I fimbrial operon of
A:Reference number: A56617; MUID:9232981
A:Accession: D56617
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-360 <JOR>
A:Cross-references: GB:M55661; NID:g145507; PIDN:AMC41417.1; PID:g145511
A:Experimental source: enterotoxigenic strain, CFA/I-57 plasmid NTP113
C:Genetics:
A:Note: sequence extracted from NCBI backbone (NCBIN:108960, NCBIP:108972)
A:Gene: cfaf
A:Genome: Plasmid
A:Superfamily: Escherichia colonizing factor antigen cfaf

```

Query Match	82.68;	Score 1558.5;	DB 2;	Length 360;
Best Local Similarity	81.28;	Pred. No. 1.2e-109;		
Matches 293;	Conservative 31;	Mismatches 36;	Indels 1;	Gaps 1

Qy	1	MNKLIFETTLPESSVLFPFAVSARKIPDEDEITNIFEGRODNESSPKHNILNNHTAYSE	60
Db	1	MNKLIFETTLPESSVGFTEFAVSADNPSEEMKNTIGHDGSSPYIINLNSTYIINING	60
Qy	61	SHLYDPMATFCLSHNHLNACPTSENPSSSSVGFENITLQFTEKRSLLKRELQIKGY	120
Db	61	SHHLXDRKSFCLSSQNLNLNACPSDAPRGATITDGENITNLTQFTEKRSLLKRELQIKGY	120
Qy	121	KOLLEFSVNCPSGLTLNSAHFNCKNNAASGASLYLYIPAGELKNLPEGGIWDATLKLRVK	180
Db	121	KOLLEFNANCPSKLALNSHFPQCNREOASGATLSLYIPAGELKNLPEGVWNAVLKLNVK	180
Qy	181	RRYSITYTYTINTIKTLDKGNLIQIWLPOKRSAPRDVLNLRPTGGGGTYIGRNSVDMCFY	240
Db	181	RRYDTYGYTYTINTIVNLTDGNLIQIWLPOKRSANARVDLNRPTGGGGTYIGRNSVDMCFY	240
Qy	241	DGYSTNSSLFETIRODNNPKRSDGKFYLKINDMOKETAYTSLLLAGKSLPTNGTSLNI	300
Db	241	DGYSTNSSLFETIRODDNSKSDGKFYLKINDDSKELVYTLSSLAGKSLPTNGQALNI	300
Qy	301	ADAASLETNMNRITAVTPELISVYLCPMGRLOJDAKENPEAGOMGNIWVTEPSSQY	360
Db	301	NFMSLETNMNRITAVTPELISVYVLCPMGRLOJDAKYKNPEAQOMGNIKITTEPSSQY	355
Qy	361	L	361

Db 360 L 360

RESULT 2

S49539

C:Cood protein precursor - Escherichia coli

C:Species: Escherichia coli

C:Date: 15-Feb-1996 #sequence_revision 01-Mar-1996 #text_change 10-Dec-1999

C:Accession: S49539

R:Froehlich, B.J.; Karakashian, A.; Melsen, L.R.; Wakefield, J.C.; Scott, J.R.

Mol. Microbiol. 12, 387-401, 1994

A:Title: Cooc and Cood are required for assembly of CSI pili.

A:Reference number: S49538; WUID:94544028

A:Accession: S49539

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-363 <PRO>

C:Cross-references: EMBL:X76908; NID:g488735; PIDD:CA54230.1; PID:g488737

C:Superfamily: Escherichia colonizing factor antigen cfaE

Query Match	50.3%;	Score 949.5;	DB 2;	Length 363;
Best Local Similarity	53.1%;	Pred. No. 7e-64;		
Matches 197; Conservative	55;	Mismatches 98;	Indels 21;	Gaps 11;

```

QY 3 KILFLETFEFSVLETPAVSADKIPGDE--SITNIF-GPR-DNRNESPKNILNNHTA 58
    | : | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 2 KRIIFL-----SIFSAVSAAGRIPEYTVGNLIRKSPAPALDMSVSPYINFTNHVAGI 57
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY 59 SESHTLYDRMFLCGLSSHNLTNGACPTSENPSSSVS-GFTNITTLQTEKRSILIKELOI 117
    | : | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 58 SLHSIVLDRIYFLCTSSSNPNYGCACPTI---GTSVQYGTITTLQTEKRSILIKRINIL 114
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY 118 KGYQLLFKSYNCSG---LTLNSHFQCKNNA-ASGASLYLYIPAGELKMLPFGGIWDT 174
    | : : : : : : : : : | : | : | : | : | : | : | : | : | : | : |
Db 115 AGNKRKPWENOSCFEINLMTVNSKSMWSCAGAHGANGTLLMLYIPAGEINTLPFGGIWEAT 174
    | : : : : : : : : : | : | : | : | : | : | : | : | : | : | : |

QY 175 LKLRYKRRYSEI-----YGYVINTIKLTLPKGNQIWLPOFKSPARDVLDLAPYGGSTY 230
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 175 LILRLHS-RYGEVSSITHYGNITVNTITVDLIDKGNIQVWLPGFHSNPRVDLTLRPIGNKY 233
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY 231 GRNSVDMCFYDGYSTNSSSLEIRFODNNPKPSDGKIFYLRKINDDTKETAYVNLSTLLAGSL 290
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 234 GSNISLDMCFYDGYSTNSDSAMVIRKQDDNPNTSSSEYNLYKIG-GREKLPYAVSLMGEKIF 292
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY 291 TPTNGTSLNINDAASLETNNRITATVAMPETISVPVLCWPGRLQDLADAKVENPERAGQIWGNT 350
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 293 YPVNGOSETINDSSVLETFNNRKYAVAMPEVNPVLCWPARLLLLNDVDNADAGQYSGOI 352
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY 351 NYTFPPSSQTL 361
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 353 YITTFPSEVNT 363
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

```

RESULT 3
S57937
Cold protein precursor - Escherichia coli
C:Species: Escherichia coli
C:Date: 13-Jan-1996 #sequence_revision 01-Mar-1996 #text_change 10-Dec-1999
C:Accession: S57937
R:Froehlich, B.J.; Karakashian, A.; Melsen, L.R.; Scott, J.R.
Submitted to the EMBL Data Library, January 1995
A:Description: The genes for Cst2 pill of enterotoxigenic Escherichia coli and their inter
A:Reference number: S57934
A:Accession: S57937
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-364 <PRO>
A:Cross-references: EMBL:Z47800; NID:9897725; PIDD:CAAB8763.1; PID:9897729
C:Genetics:
A:Start codon: TTG
A:Superfamily: Escherichia colonizing factor antigen cfaE

Query Match	49.38;	Score 930.5;	DB 2;	Length 364;
Best Local Similarity	48.48;	Pred. No. 1.9e-62;		
Matches 177;	Conservative 64;	Mismatches 118;	Indels 7;	Gaps 3;

QY	1	MNKLIFLFTLFESSVLETFNVA	SADKIPGDSI	ITNIFGPRDRN--	ESOPKINILNHNITAY	58
Db	1	KKKIIFFVLSMFLKQYVYGQSH	HTNVEGSI	NKTEESICP	IDRASAASYPAHYITHEHVAGY	60
QY	59	SESHFLVDRMFLC	LSHNTLNGACPT	SENPSSSSVG	ETNITLOFTFKRSILKRELOK	118
Db	61	NKDSHFLDRMFLC	MSSTDAKSCAP	TGENSEKS--	QGGETNKLIFTEKKSILARTLNK	118
QY	119	GYKOLLEKRSVNC--	PGGLILNSAHFNCKNNA	SGASLYIT	ITAGELKNI	175
Db	119	GYKRFLEYESDRC	HIYVDKMNLSHTVC	SVGSTRFGVDFT	LYIFOGELIDGLTGTGIEWATL	178
QY	176	KLRKRRSEIYGT	YITNITIKLIDKGNIO	IMWPOFKSDAR	VDNLNRP	235
Db	179	ELRRKRRHDIYHG	YKVNITVLDLTKGNIO	WTPPKFHS	DRIDLNLNRP	238
QY	236	DMCFYDGYSTN	SSSLSEIRFODNNK	SDGKEFLRK	INDTKEIAYT	295
Db	239	EMCLINDGYSTH	SQSIEMRFPQDSOT	GTGNNEVNLIT	GTGPIKLKLPYK	298
QY	296	TSLNIDAAASLET	WNKRITAYTMPE	ISVPVLCW	GRLOLDKAYEN	355
Db	299	EAFINIDTSS	FLIMWNRIKYS	LSLOISIPAL	CPWNPANLTFE	358
QY	356	PS	SOTL	361		
Db	359	PS	SSSL	364		

RESULT 4
AE0541
Probable fimbrail protein tcfD [imported] - *Salmonella enterica* subsp. *enterica* serovar
C:Species: *Salmonella enterica* subsp. *enterica* serovar Typh
A:Note: this species has also been called *Salmonella typhi*
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 09-Nov-2001
C:Accession: AE0541
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church
th, T.; Connetton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farr
S.; Moule, S.; O'Garra, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,
A.:Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* se
A:Reference number: AB0502; PMID:11677608
A:Accession: AE0541
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-359 <PAR>
C:Cross-references: GB:AL513382; PIDN:CAD08773.1; PID:g16501589; GSPDB:GN00176
C:Genetics:
;Gene: tcfD

	Query Match	14.6%	Score 275.5;	DB 2;	Length 359;
	Best Local Similarity	29.7%	Pred. 3.1e-13;		
	Matches	90;	Conservative	48;	Mismatches 130; Indels 35; Gaps
QY	70 FLCLSHNTLWGACPTSENPSSSSSGEINTITFTFKRLIRKELQIKGKOLLERSVN	129	:::::	:	: :
Db	79 WVCNSNRRNGEACEEHILVWVFAGAYSKIRLRFRQISHAETLL-----ILLGSVR	131	:::::	:	: :
QY	130 --CGSLTLNSAHENCKNNAA--GASLYLYIPDAGEKLKLPFGGIWDATLKL-RVKRRY	183	:::::	:	: :
Db	132 DACTGV-----IMNNAACQMGSRSLKRIRPSEEIAKIPISGTWKALIVLDYDLQMG	183	:::::	:	: :
QY	184 SETGYTINTITIKLTDP-KGNIDWLRF-QKSARVDLMLRTPGGTTYIGRSVMCFY	240	:::::	:	: :
Db	184 DDPLCTSTDTLLTWLVTHFAENAIYPFGCTAPRPADLLHMMNNSQMGRANLMCLX	243	:::::	:	: :

```

Oy      241 DGYSTNSSLEFRQDNNPKSDGKFYLKINDPDKREIAYVTLISLLAGSLFFPTNGTSLNI   300
           || : |:::~::~|| ::::| | | :||::| | | ~:::
Db      244 DG-GYKAASLDMKEIGSKSGTGFOVIKSADVT--IDYAVSMMNYGGSIPTVRGEFSL   300
                                           : | | : | | : | |
Oy      301 ADAASLENNMRITAVTPPEISVPILCMWGRGLDLDAK---VENPEACGYMGNINVTFTPS   357
                                           : | | | | | : | | : | |
Db      301 DNVDKAAAR-----PVVLPGGRQAVRCVPPLVLTLTGPFNIREKRSSGEYGCTLVTTMLMG   355
                                           : | |
Oy      358 SQTF 360
           : ||
Db      356 TQT 358

RESULT    5
G90975
Probable factor [imported] - Escherichia coli (strain O157:H7, substrain RIMD 0509952)
C:Species: Escherichia coli
C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C:Accession: G90975
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.-G.
gasawara, N.; Yasunaga, T.; Kunihara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
A>Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic
A:Reference numbers: A99629; MUID:21156231; PMID:11258796
A:Accession: G90975
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1335 <HAY>
A:Cross-references: GB:BA000007; PTDN:BAB36198.1; PTD:g13162243; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RIMD 0509952
C:Genetics:
C:Gene: ECs2775
```

Query Match	6.4%	Score 120;	DB 2;	Length 1335;
Best Local Similarity	20.4%	Pred. No. 0.85;		
Matches 80;	Conservative 66;	Mismatches 156;	Indels 90;	Gaps 15

[illegible]

```

RESULT      6
E85822
probable Invasin Z3135 [imported] - Escherichia coli (strain O157:H7, substrain EDL933)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001

```

C:Accession: E85822
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
Miller, L.; Grobbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouzis, K.; Apoda
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: E85822
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-2660 <STO>
A:Cross-references: GB:AE005174; NID:912516151; PIDN:AAG57041.1; GSPDB:GN00145; UWGP:
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: Z3135

Query Match	6.4%;	Score 120;	DB 2;	Length 2660;
Best Local Similarity	20.4%;	Pred. NO. 2.1;		
Matches	80;	Conservative	66;	Mismatches 156;
			Indels	90;
			Gaps	15;

```

Oy 13 SSVLETFVAVSADKJPGGEESTJTFNFGRRDRRESSPKHINILHITAY - - - - -SESHITLYDDMT 69
Db 737 SAKIATISASINGVLANENANVTAVSVNVADEGS - - - - -NPINDTIVTFAYLSGATISFNQON 793

Oy 70 FICLSHNTINGACPTSPENSSSVSGEITITLOFTEKRSLI - - - - -K 112
Db 794 - - - - -TATDVNGLA - FFDLKSSKQEDONTJEVTELENVKOTLLVSPFGDSSTQVODLOKSK 848

Oy 113 RELOIKGY - KOLLEKFSVNCPSGLTLSAHBNCKNKNASGASLYLYIPAGELKMLPFEGIM 171
Db 849 NEVVADENDSATMAYATRADKAGMLNDVKTFFVNSAAKLSQTEVNSHD - - - - -GIA 901

Oy 172 DATLKLAKRRYSSTVYTYINTITIKLJDKGNQIQLPQFKDARVDLMLRPTGGGYIG 231
Db 902 TATLT - - - - -SLKNDYVTVTASVSSGSOANOQVITGDSOTATLTLSV - PSGDITV - - 951

Oy 232 RNSVDMCFYDGYSTNSLSLE - - - - -IRFODNPFKSDKFLRKINDTTEIATVLTLLAG 287
Db 952 - - - - -TNTPLIMTATLODKKNPLAKDEITFSVPND - - - - -VASRFSISNG 994

Oy 288 KSLTPTNGTSL - - - - -NIADASLETNMNRITAYMEISVPVC 327
Db 995 KGMDSNGOTAIASLTGLTACTHMTARIANSVNSDQPTFVADRDRAVVVLQTSKAEII 1054

Oy 328 WPG - - - - -RIQDQAKVENPEAGGYMGNINVTTPS 357
Db 1055 GNGVDFTTLTATYKDP - FDNVNVNLSVFRPTS 1085

```

RESULT 7

alpha-agglutinin - yeast (*Saccharomyces cerevisiae*)
S22835

N:Alternate names: 22k glycoprotein; protein J118; protein YJR004c
C:Species: *Saccharomyces cerevisiae*
C:Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 29-Oct-1999
C:Accession: S22835; S51229; A32822; S55192; S57019
R:Hauser, K.; Tanner, W.
FEBS Lett. 255: 290-294, 1989

A:Title: Purification of the inducible alpha-agglutinin of *S. cerevisiae* and molecular
A:Accession: S22835; MUID:90005993
A:Reference number: 522835

A:Molecule type: DNA
A:Residues: 1-650 <HAU>
A:Cross-references: EMBL:X16861; NID:g3352; PIDD:CAA34752.1; PID:g3353

A:Accession: S51229

A:Molecule type: protein
A:Residues: 20-24 <HA2>
R:Lipke, P. N.; Wojciechowicz, D.; Kurjan, J.
Mol. Cell. Biol. 9: 3155-3165, 1989

A:Title: AG-alpha-1 is the structural gene for the *Saccharomyces cerevisiae* alpha-agg
A:Reference number: A32822; MUID:90014768
A:Accession: A32822

A:Molecule type: DNA

A:Residues: 1-448, 'P', 450-555, 'E', 557-580, 'L', 582-650 <LIP>
A:Cross-references: GB:M28164; NID:g171041; PIDN:AAA34417.1; PID:g171044
R:de Haan, M.; Smits, P.H.M.; Grivell, L.A.
submitted to the EMBL Data Library, May 1995
A:Reference number: S55183
A:Accession: S55192
A:Molecule type: DNA
A:Residues: 1-650 <DEH>
A:Cross-references: EMBL:X87611; NID:g854567; PIDN:CAA60926.1; PID:g854577
R:de Haan, M.; Grivell, L.A.; Smits, P.H.M.
submitted to the Protein Sequence Database, September 1995
A:Reference number: S56771
A:Accession: S57019
A:Molecule type: DNA
A:Residues: 1-650 <ZAG>
A:Cross-references: EMBL:Z49504; NID:g1015625; PIDN:CAA89526.1; PID:g1015626; MIPS:YJR00
C:Genetics:
A:Gene: SGD:SAG1; AGAL1; AGALPHA1
A:Cross-references: SGD:S0003764; MIPS:YJR004C
A:Map position: 10R
C:Keywords: glycoprotein

Query Match 6.0%; Score 114; DB 2; Length 650;
Best Local Similarity 20.9%; Pred. No. 0.94;
Matches 98; Conservative 52; Mismatches 147; Indels 172; Gaps 23;

OY 4 ILEIFLFPSSVL-----FFFA-----VSADKIRG-----DESTINIFGRDRNE--- 43
DB 8 IIMLEFLALASAININDIFFSLEIPPLTANKOPDOGWATFSPDIADASSIEGGEFTL 67
OY 44 SSP---KHNILNNHITA-----YSESHITLYDRMTFLC-----LSHNTLN 80
DB 68 SMPHYVRIKILNNSQATISLADTEAFKCYVSGQAAYILENTFTCTAQNDSISNTID 127
OY 81 GACPTSENPSSSVSEETNTITLOFTEKRSIKRELQIKGKQLFFKSVNC-PEGGLNLSA 139
DB 128 GSTTFSLNPSDGGSSYEEL-----ENAKFKSGPMILVKGNDMSDVNPDPAFTENVF 182
OY 140 HFNCNNAASGASLYL--YIPAGELKNLPRGGIMDATIKLRVRRSEYGTITNTIK 197
DB 183 HSGRSTGYSGFSEYHLGMCPCNGYF---LGG---TEKI-----DDSSNNVNDL- 225
OY 198 LVDKNGIQI-----MLPQKSDARVDNLPRTGGTYIGRN---SVDMCFYDG--- 242
DB 226 --DCSSVOYSSNDPMDWMPQSYNDTNDV-----TCFESNLMITDELXLYGEML 275
OY 243 -----YSTNSSLE--IRPDNN--PKSDGKF 265
DB 276 WVNALQSLPANVNTIDHALEFOYTCDDTANTYATQFSTREFIYVOGRNLGTASAKSS 335
OY 266 YLRKINDPRKEI--AYTLS-----LLAKSLPLPTNGTSLNTADA 304
DB 336 FISTTTTDLTSLNTSAISGSIISTVETGNNRTTSEVISHSVTTSTKLSPTATSTLTAAQS 395
OY 305 SLETNMN-----RITAVTPEISVPVLCMPGRL 332
DB 396 IYSTDNTITVGTIDHTTSEVISDVETISRETASTVVAAPSTGTGWMGAM 444

RESULT 8
D64962 Probable membrane protein b1978 - Escherichia coli
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 02-Feb-2001
C:Accession: D64962
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; CD
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: D64962
A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA
A:Residues: 1-2383 <BLAT>
A:Cross-references: GB:A6000289; GB:U00096; NID:g1788285; PIDN:AAC75042.1; PID:g17882
A:Experimental source: strain K-12, substrain MG1655
C:Keywords: nucleotide binding; P-loop; transmembrane protein
F:54-70/Domain: transmembrane #status predicted <TM>
F:1564-1571/Region: nucleotide-binding motif A (P-loop)

Query Match 5.8%; Score 109; DB 2; Length 2383;
Best Local Similarity 20.0%; Pred. No. 12;
Matches 78; Conservative 65; Mismatches 157; Indels 90; Gaps 15;

OY 13 SSVLTFEAVSADKIPDESDITNIFGRDRNESSPKNIILNNHTAY---SESHITLYDRMT 69
DB 762 SATIATLSSNNQVLANENAAATVSVNVADEGS---NPINDHVTVPVAVLSGATSTNNON 818
OY 70 FLCLSSHNTLNGACPTSENPSSSVSEETNTITLOFTEKRSI-----K 112
DB 819 ---TAKTDVNGLA--TFDLKSSKQEDNTVEVLENGVKQTLIVSEVGDSTAQVDLQSK 873
OY 113 RELQITGYKOL-LFKSVNCPSCGLTNSAHFNCKNAASGASLYIYPAGELKNLPRGGIW 171
DB 874 NEVVADGNDSVTATVTRAKGNLNDVAVTFVNSAEAKLSQTEVNSHD-----GIA 926
OY 172 DATLKLRRVRRSEYGTITNTIKLTDKGNQIWLPPQKSDARVDNLPRTGGTYIG 231
DB 927 TATLT-----SLKNGDYRVTRASVSSGQANQVAFIDQSTALITLSV-FSGDITV- 976
OY 232 RNSVDMCFYDGYSTNSS---SLEIRPDNNPKSDGKFYLRIKINDPKEIAYTSLIAG 287
DB 977 -----TNTAPQYMTATLQDKNGNPLDKKEITFEVSPND---VASKFSISNGG 1019
OY 288 KSLTPNGTSL-----NTADAASLETMMNRTTAVTMEISVPVLC 327
DB 1020 KGMTDSNGVALISLTGTLGTHIMARLANSVNSDQPMTEFVADKRAVAVVLTQSKAEIT 1079
OY 328 WPG--RIQLDAKVENPEAGQYMGNIWTF 355
DB 1080 GNCVDETTLTATVYKDP-SNHPAVAGITVNT 1108

RESULT 9
S55097
Probable membrane protein YMR215w - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein YMR261.09
C:Species: Saccharomyces cerevisiae
C:Date: 08-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 05-Nov-1999
R:Dedman, K.; Brown, D.; Bowman, S.
submitted to the EMBL Data Library, June 1995
A:Reference number: S55089
A:Accession: S55097
A:Molecule type: DNA
A:Residues: 1-524 <DED>
A:Cross-references: EMBL:Z49809; NID:g854459; PID:g854467; GSPDB:GN00013; MIPS:YMR215
A:Experimental source: strain Ab972
C:Genetics:
A:Gene: MIPS:YMR215w
A:Map position: 13R
C:Superfamily: glycopospholipid-anchored surface glycoprotein GAS1
C:Keywords: transmembrane protein
F:7-23/Domain: transmembrane #status predicted <TM>

Query Match 5.7%; Score 108; DB 2; Length 524;
Best Local Similarity 24.7%; Pred. No. 2;
Matches 69; Conservative 42; Mismatches 96; Indels 72; Gaps 18;

OY 119 GYKOL-LFKSVNCPSCGLTNSAHFNCKNAASGASLYIYPAGELKNLPRGGIMDA 173
DB 264 GYDKLNSTFEDAVIP---LIFSEYGCNKNTPRTEDEVSEGLY---GGLKNVSGGL--- 313

QY 174 TLKRVKRRSEYGYTITITIKLFDKGNIOIWLPOFKSD-ARVDLNLK-----PTGGCT 228
 Db 314 -----VEETEEANNYGL-----VKLDDSGSL-----TYKDFEYNLSQNLNVLPTTKES 339
 QY 229 YIGRNSVDMC-----FYDGYSTNSSSLE-----TRFODNNPKSDGK----- 264
 Db 360 EISDSIYKODNSAINTIYSGFCTNNFTLPSQAEIAIAMEYGVNGTNT-CKILITDYAVP 418
 QY 265 ----FLIKRINDTKELIAYTILSLLAGKSLTPNGTSLNTADASLETNNR-ITAVTMP 319
 Db 419 TPENYTIKNNKDDT--ISATISYDKA-NSLNEIDVTATYVAKSASTSSQSSRSLSSTSP 475
 QY 320 EISVPILCMGRLQLDKAVENPEAGYMGNI-VTFPPS 357
 Db 476 SSTGSSSSSTG-----SSASSSSKSGVGNIVVSEFSQS 510

RESULT 10

B64635
 toxin-like outer membrane protein HP0922 - Helicobacter pylori (strain 26695)
 C:Species: Helicobacter pylori
 C:Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999
 C:Accession: B64635
 R:Tomb, J.F.; White, O.; Kervatage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenney, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.; Nature 388, 539-547, 1997
 A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.
 A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
 A:Reference number: A64520; MUID:97394467
 A:Accession: B64635
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-2529 <TON>
 A:Cross-references: GB:AE000602; GB:AE000511; NID:g2314060; PIDN:AAD07969.1; PID:g231406

Query Match

Best Local Similarity 5.7%; Score 108; DB 2; Length 2529;
 Matches 84; Conservative 50; Mismatches 122; Indels 118; Gaps 22;

QY 85 TSENPSVSSVSGETITLOFTEKRSLLI-----KRELOIKGYKQLLFK---SYNCP 131
 Db 977 TSNFNATITQLGNTNFTL---SSQILNFGDITLONNANITIGNKSQAFAKSLIIDDN 1033
 QY 132 SGLTL-NSAHFNCNKNA--SGASLYLYTPA-GEKRLPEGGIMDATLKLVRKRYSEY 187
 Db 1034 SNLSLNDQSVLANMNTSAFNQASLNIYNGSQATFNSLPFNG---GTLNLASSKLNASN 1090
 QY 188 GTTITITIKLTP-----KGNITQIWLPOFKSDARVDLNLKPTGGCTYIGRNSVD 236
 Db 1091 ASFSNNTTITLNDQSVLASNTSSLNANI-----NFGQASQADP---GONTITIDTASFN 1140
 QY 237 MCFYDGYSTNSSSL---EIRFODNNPK-----SDGKFTYRKIN---DDT 274
 Db 1141 --PDSASSLNFNLTANGALNFGITPSSLKALMSVSGCVLGNNGDINLSINDIINDIT 1198
 QY 275 KEIAYTILSLLAGKSLTPNGTSLNTADASLETNNR-ITAVTMP 319
 Db 1199 KSVTY--NILMAQGITIGSANGYEKILFYGKIKQIATYSDNNNTQGTWTFINPLNSQI 1256
 QY 322 -----SVFVL-----CMFGRQLDKAVENP-----EAGQY--M 347
 Db 1257 IOESIKNGDLTTEVLNPNNSASNTIFNIAPELYNQASKONPTGYSYSDNOAGTYIYL 1316
 QY 348 GNINVTFTP-SSOT 360
 Db 1317 SNIKGLFTPKGSOT 1330

RESULT 11

A64904
 probable fimbrial protein b1502 - Escherichia coli

C:Species: Escherichia coli
 C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 29-Sep-1999
 C:Accession: A64904
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
 .A.; Rose, D.J.; Mau, B.; Shao, Y.
 Science 277, 1453-1462, 1997
 A:Title: The complete genome sequence of Escherichia coli K-12.
 A:Reference number: A64720; MUID:97426617
 A:Accession: A64904
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-304 <BLAT>
 A:Cross-references: GB:AE000247; GB:U00096; NID:g1787773; PIDN:AACT4575.1; PID:g17877
 A:Experimental source: strain K-12, substrain MG1655
 C:Superfamily: fimbrial protein fimH
 C:Keywords: fimbria

Query Match

Best Local Similarity 5.7%; Score 107; DB 2; Length 304;
 Matches 75; Conservative 51; Mismatches 125; Indels 106; Gaps 19;

QY 3 KILF-ITLFFSVLFTFAVSADKIPGDESI---TNIEGPRDRNNSPKHNT---LNNH 54
 Db 8 KVLFGIYLLMAKRVAFSCNVN---GGSSIGAGTISVYNLD-PVIGPQNLVLDLSQH 63
 QY 55 ITAVSESHLYDRMTFLCLSSH-NTLNGACPTSENDSSSVSGETITLOFTEKRSLLIKR 113
 Db 64 ISCMNDYGGWYD-----TDHINLVQGG-----SAFAG----- 89
 QY 114 ELQIKYK-QLLKSVNCPGCLTLNSAHFNCNKNAASGASLYLYIRAGELKNLPFGCIWD 172
 Db 90 --SLQSTKGSLYNNVNTYPPPLTTNTNVDIGKTPMPLKLYI-----TPVGAAG 140
 QY 173 ATK-----LRVRRYSEYGT-----YTITITIK-----LTDKGNIOIWLPOFK 212
 Db 141 VVIKAGEVIAIIMYKATIGSGNPRFTNIIISNNNVVPTGGCTYDSNIVVDLPDF 200
 QY 213 SDARVDLNLKPTGGCTYIGRNSVDMCFYDGYSTNSSSLETIRFODNNPKSDGKFTYLRKIND 272
 Db 201 GSAEIPF-----GVYCS--SEOKLSEFLSGAATDSSRQV-FANTAP-----D 239
 QY 273 DTEKIAVITISLLAGK-----SLPTNGTSLNTADASLETNNRITAVTMP 319
 Db 240 ATRASGVGVTLMRKILATGENVSLGTVNKSVPVLGSLATYQGTQGNKVSAGTVQSV 296

RESULT 12

C90892
 probable adhesin [Imported] - Escherichia coli (strain 0157:H7, substrain RIMD 050995

C:Species: Escherichia coli
 C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
 C:Accession: C90892
 R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.
 gaeawara, N.; Yasunaga, T.; Kihara, S.; Shiba, T.; Hattori, M.; Shingawa, H.
 DNA Res. 8, 11-22, 2001
 A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli 0157:H7 and g
 A:Reference number: A99629; MUID:21156231; PMID:11258796
 A:Accession: C90892
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-304 <HAY>
 A:Cross-references: GB:BA000007; PIDN:BA035530.1; PID:g13361573; GSPDB:GN00154
 A:Experimental source: strain 0157:H7, substrain RIMD 0509952
 C:Genetics:
 A:Gene: ECS2107
 C:Superfamily: fimbrial protein fimH

Query Match

Best Local Similarity 5.7%; Score 107; DB 2; Length 304;
 Matches 75; Conservative 51; Mismatches 125; Indels 106; Gaps 19;

```

QY 3 KILF IFLFESSVLETFVAVSADKIPGDESI---TTFEGRRDNRESSPKHHI---LNNH 54
Dh 8 KVLFGIYLLMAKGFARFASCND---GGSSIGAGTTSYYVIND -PVIGPQNLYVDDLSQH 63
QY 55 ITAVSEHTLLDRLMFLCLSSH-NLNGACPTSENPSSSVSGEINITLQFTEKSLIKR 113
Dh 64 ISCWMDYGGWMD---TDHINLVQG---SARAG----- 89
QY 114 ELQIKGYR-OLLEFKSVNCPGILTLLNSAFHNCNKNAASGASLYLIPAGELKNLPFGIAMD 172
Dh 90 --SLQSGSLYMNWNVTPFPLTTNTNVLDIGDKTPMLPKLYI-----TPVGAAG 140
QY 173 ATKK-----LWKKRYSETYGF-----YTINTTK-----LTKGNIQILPQFK 212
Dh 141 VVIRKGEVIARIHKYKATLLSGNPNFTWNTISNNSVYMPFGCTYVSRYNTVNLPDFP 200
QY 213 SDARVDLMLRPTGGGTLYIGRNSVDMCFYDGYSTNSSLERIFODNNPKSDGKFYLKIND 272
Dh 201 GSAEPL-----GVYCS-SFOKLSFYLSGTGTTDSARQY-FANAP-----D 239
QY 273 DTKELAYTLISLLLAK-----SLTPNGSLNIAAASLETWMNRITAVTMEI 321
Dh 240 ATKASGVGSLARNKIIATGENVSLGTVNKSKYPLGSLATYGGQTGNKSVSGATVOSV 296

```

RESULT 13
F85725
probable adhesin, fimb type protein Z2206 [imported] - Escherichia coli (strain O157:H7,
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: F85725
R:Ferna, N.T.; Pinkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Hiller, L.; Grobbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamoultis, K.; Apodaca
Native 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: F85725
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-304 <STO>
A:Cross-references: GB:AE005174; NTD:G12515169; PION:ANG56266.1; GSPDB:GN00145; UWGP:Z22206
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: Z2206
C:Superfamily: fimbrial protein fimb

Query Match	5.78;	Score 107;	DB 2;	Length 304;
Best Local Similarity	21.08;	Pred. No. 1.2;		
Matches	75;	Conservative	51;	Mismatches 125; Indels 106; Gaps 19

```

QY 3 KILF-IFLFFSVLEFPAVSADKIPGDESI---TNTFEGPDRRESSPKHII---LNHH 54
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 8 KVLBEIYLLMAAGKFAFASCVND---GGSSIGAGTISYVND- PVIQPGNULVYDLSQH 63

QY 55 ITAYSESHLIDRMTFLCLSSH-NTLNGACPPSENPPSSVSGETNITLOFTEKRSILKR 113
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 64 ISCVMDYGGWYD---TDHINLVQG---SAFAG----- 89

QY 114 ELOIINGYR-OLLEFKSVNCPGILTNSAHFNCKNNAASGASYLYIYPAGELKNLPGGIWD 172
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 90 --SIQSYGSLYMNWVYLPPLTTNTNVLDIGDKTPMPLPKLYI-----TPGGAAG 140

QY 173 ATLK-----LKYKRYSETYGT-----YTINITIK-----LTDKGNIOIWLPOFK 212
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 141 VVIAGEVIARIHMXYKIATLFGSGNPRNFTWNIISNNSVYMPGTGCTYVSRYNTVNLPPFR 200

QY 213 SDARVDMLRPETGGCTYIGRNSVDMCIFYDGYSTNSSLIEIRFOODNPPKSDGKFYLRKIND 272
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 201 GSAEPL-----GVYCS-SBQKLSFYLSGTTDSANQV-FANNAF-----D 239

QY 273 DTKELATYLSLLLAGK-----SLTPNGTSLNIAAASLETWMNKTAVATMEI 321

```

Db 240 ATKAGVGVSLMRNGILATGENTVSLGTNKSXPVLGSLATYGGTGNKVSAGTVO5V 296

RESULT 14

A38109

autolysin - Enterococcus faecalis

C:Species: Enterococcus faecalis

C:Date: 24-Jul-1992 #sequence_revision 24-Jul-1992 #text_change 15-Oct-1999

C:Accession: A38109

R:Beliveau, C.; Potvin, C.; Trudel, J.; Asselin, A.; Bellemare, G.

A:Title: Cloning, sequencing, and expression in *Escherichia coli* of a *Streptococcus* f

A:Reference number: A38109; MUID:91358349

A:Accession: A38109

A:Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-671 <BRL>

A:Cross-references: GB:M58002; NID:g153658; PIDN:AAA67325.1; PID:g829194

	Query March	5.6%:	Score 106;	DB 2:	length 671;	
	Pred Local Similarity	23.6%:	Pred No. 3.9;			
	Matches 73; Conservative	39;	Mismatches 103;	Indels 94;	Gaps 17;	
QY	41 RNESPKEHNI-LNNHITAYSESHLYDRMTFLCSNHTLNGACPTSENPSSSV-----	94				
Db	309 RYATDPSTNAKLNNVTIAY--NLTYDTPSSGCGTGGGTVPBGTSNNQSTNTITYTVK	366				
QY	95 SCGEI--NITLOP----TEKRSL--IKRELOIKGYKQLLFKSYNCPSGLTNSAHENCUKN	146				
Db	367 SGGDTLNKIAAOYGVSVALNRSMNGISGDILFVGOKLIYKKGA---SGMTGSGSNGGSNNN	423				
QY	147 AASGASLLVIYPAGELKMLPGGINDATFKLRVKRYSSETGYTIINTLIKDKGNIOI	206				
Db	424 -QSNGNTIYYTKSGDTLN-----KIAAQIG-----VTV-----ANLRS	455				
QY	207 WLPQKSDARYDLNL-----RPFGGQTYIGRNSVDMCFEYGSTNSSLSLEIFODNN	258				
Db	456 W-----NISGDILFVGOKLIYKKOTSGNT-----GSSNGGS-----NNN	491				
QY	259 PKSDCKFLRKINDDTKELA--YTLS-----LLAGKSLEPTMGTSLINIDA	303				
Db	492 QSGTTYTYTIKSGDTLNKIAOYGVSVALNRSMNGISGDILFAGOKIIVYKKGTSGNTGGS	551				
QY	304 ASLETNMNR	312				
Db	552 SNGSGSNNO	560				

AC2507
hypothetical protein all17235 [imported] - Anabaena sp. (strain PCC 7120) plasmid
C:Species: Anabaena sp.
A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002
C:Accession: AC2507
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriiguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium
A:Accession number: AB1807; MIMD:21595285; PMID:11759840
A:Accession: AC2507
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-843 <KUR>
A:Cross-references: GB:BA000020, PIDN:BA078319.1; PID:917135773; GSPDB:GN00180
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all17235
A:Genome: plasmid
Query Match . . . 5.6% Score 105.5; DB 2; Length 843;

Best Local Similarity 18.9%; Pred. No. 5.8;
Matches 71; Conservative 67; Mismatches 127; Indels 111; Gaps 18;

```

0Y 77 NTLNACCTSEN-----PSSSVGETNITLQ-----PT--EKSLKREL 115
   | : : : |
Db 297 NILDGSLVFTQNHGFKGAVKIDAOISLTIQSSMLAISLYTSMFGCTPPEESTIQLDVKK 356
   | : : : |
0Y 116 QIKGYKOLLFRKSVNCPGSL-TLNSAHFN-----CNKAASGASLYLYIP-----AG 160
   | : : : |
Db 357 TIOGQIATTFETNAPSOLLIINSLSKISODPDSVYANPDOLGQINTFVSYSCKGGDIAG 416
   | : : : |
0Y 161 ELKNIPEG-GIMD-----ATLKLRYKRRKSEYIYTYITITIKLIDKGNIQIIMLP 209
   | : : : |
Db 417 KINNIILDSVFTNVAASGAGCNPLELENLIKNGGASLSSTIRSGOGVNV----- 471
   | : : : |
0Y 210 QKSDARVDLN-----LRPT-----GGGYIGRNSVDMCYDOSTNNSGLE----- 251
   | : : : |
Db 412 -FKSQNIDIGGQALRFPSTITSTTGHGNGNIDINTLNLISNGGGISSTLSAGKA 530
   | : : : |
0Y 252 -----IRFODNNPK-SDGKFYLRKINDPRKEIAYTLLSLACKS----- 289
   | : : : |
Db 531 GNISINSNSINAVGTNINSNSPFIINSNELL--VDPNLOKLLYLRQPELLIOAGNIEL 588
   | : : : |
0Y 290 -----LTPNGTSLN-----IADAASLETNNRITAYTMPEISVYPLCWP-GRLOLDAYE 339
   | : : : |
Db 589 NTDIINISNGCLINARNEGVDACNIRISANTININSGEVENATTTIGEGGNIILNSR-- 646
   | : : : |
0Y 340 NPEAGQVYMGCNINVTFT 355
   | : : |
Db 647 -----NLFNLSNRITAT 658

```

Search completed: July 1, 2002, 11:24:57
Job time: 249 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 1, 2002, 11:23:44 ; Search time 13.41 Seconds
(without alignments)

1042.338 Million cell updates/sec

Title: US-09-839-894-10

Perfect score: 1886

Sequence: 1 MNKILFIFLTFSSVLETFFA.....EAGQYMGNINVTTPSSQTL 361

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1558.5	82.6	360	1	CPAE_ECOLI
2	114	6.0	650	1	SAG1_YEAST
3	108	5.7	524	1	GAS3_YEAST
4	107	5.7	304	1	YDEQ_ECOLI
5	106	5.6	671	1	ALYS_ENTFA
6	99	5.2	1849	1	IG44_HAEIN
7	98.5	5.2	608	1	YD93_METUA
8	98	5.2	758	1	SP21_YEAST
9	98	5.2	1157	1	NI33_YEAST
10	95.5	5.1	713	1	CD62_PAEWA
11	95	5.0	413	1	RF1M_YEAST
12	95	5.0	2334	1	WAPA_BACSU
13	94.5	5.0	959	1	NI00_YEAST
14	94	5.0	471	1	LEU2_BUCRP
15	93.5	5.0	412	1	THRG_SHEEP
16	93	4.9	1167	1	CLAA_BACTU
17	92.5	4.9	320	1	CYF_CVACA
18	92.5	4.9	1061	1	AGM1_CANAL
19	92.5	4.9	1061	1	OAR_MYXXA
20	92.5	4.9	2278	1	FABL_YEAST
21	91.5	4.9	869	1	CFPC_ECOLI
22	91.5	4.9	1783	1	Y468_MYCGE
23	90.5	4.8	309	1	ITRB_CLOPE
24	90.5	4.8	863	1	MCW4_XENLA
25	90.5	4.8	1019	1	ENRK_HUMAN
26	90.5	4.8	1328	1	HUS2_SCHPO
27	90.5	4.8	2193	1	POG6_HE71M
28	89.5	4.7	1113	1	WGA2_YEAST
29	89	4.7	465	1	ANT3_MOUSE
30	88.5	4.7	455	1	MURF_BUCAI
31	88	4.7	1394	1	HAP_HAEIN
32	87.5	4.6	468	1	VGIC_HSVB
33	87.5	4.6	792	1	UBPA_YEAST

34	87	4.6	711	1	ETP2_SPVKA
35	87	4.6	1290	1	BXCL_CLOBO
36	86.5	4.6	411	1	THRG_BOVIN
37	86.5	4.6	1002	1	SPHR_AMEPV
38	86.5	4.6	1273	1	WEI1_YEAST
39	86.5	4.6	1540	1	RM1_DROME
40	86.5	4.6	1769	1	YJG9_YEAST
41	86	4.6	368	1	YB11_YEAST
42	86	4.6	493	1	FLIC_SALPA
43	86	4.6	828	1	MRC_KLEPN
44	86	4.6	1211	1	BUN2_DROME
45	86	4.6	1304	1	CD45_HUMAN

ALIGNMENTS

RESULT	1	STANDARD	PRT	360 AA.
CPAE_ECOLI				
ID	CPAE_ECOLI			
AC	P25734			
DT	01-MAY-1992 (Rel. 22, Created)			
DF	01-MAY-1992 (Rel. 22, Last sequence update)			
DT	01-AUG-1992 (Rel. 23, Last annotation update)			
DE	CFA/I fimbrial subunit E (Colonization factor antigen I subunit E).			
GN	CPAE.			
OS	Escherichia coli.			
OG	Plasmid NTP513.			
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;			
OC	Escherichia.			
OX	NCBI_TaxID=562;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=92329981; PubMed=1352712;			
RA	Jordi B.J.A.M., Willshaw G.A., van der Zeijst B.A.M., Gaastria W.;			
RT	"The complete nucleotide sequence of region I of the CFA/I fimbrial			
RT	operon of human enterotoxigenic Escherichia coli.";			
RL	DNA Seq. 2:257-263(1992).			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation-			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; M55661; AAC41417.1; -			
KW	Antigen; Fimbria; Plasmid.			
SQ	SEQUENCE 360 AA; 39903 MW; 69150963ABE99CE CRC64;			

Query Match 82.6%; Score 1558.5; DB 1; Length 360;
Best local similarity 81.2%; Pred. No. 1.4e-112;
Matches 293; Conservative 31; Mismatches 36; Indels 1; Gaps 1;

QY	1	MNKLIFLTFSSVLETFFAVADKIPGDESTINIRPRDRNSSPKHNLNHNITAYSE 60
DB	1	MNKLIFLTFSSVLETFFAVADKIPGDESTINIRPRDRNSSPKHNLNHNITAYSE 60
QY	61	SHLLYDRMPFLCLSSHTNLGACPTSENPSSSVSGETNTTLQFTKRSILIKRELQIKGY 120
DB	61	SHLLYDRMPSFLCLSSQNTLNGACPSDAPGTATIDGTNTTLQFTKRSILIKRELQIKGY 120

OY 121 KOLFFSVNCPGLTNSAHFNCKNNAAGSALYLYIPAGELKNLDPGGIMDTATLKLRYK 180
 DB 121 KOLFFSVNCPGLTNSAHFNCKNNAAGSALYLYIPAGELKNLDPGGIMDTATLKLRYK 180
 OY 181 RRSSEYGTITITITKLDKGNIOIMLPQFKSDARVDLNRPTGGTYIGRNSVDMCFY 240
 DB 181 RRSSEYGTITITITKLDKGNIOIMLPQFKSDARVDLNRPTGGTYIGRNSVDMCFY 240
 OY 241 DGYSTNSSLEIRFQDNPNPDKFYLKRIINDTKEIAYTLISLLAGKSLPTNGTSLNT 300
 DB 241 DGYSTNSSLEIRFQDNPNPDKFYLKRIINDTKEIAYTLISLLAGKSLPTNGTSLNT 300
 OY 301 ADAASLETNMRITAYTMEISVPLCWPGRIOLDKAVENPEAGQYMGINVTFTPSSTQ 360
 DB 301 ADAASLETNMRITAYTMEISVPLCWPGRIOLDKAVENPEAGQYMGINVTFTPSSTQ 360
 OY 361 L 361
 DB 360 L 360

RESULT 2
 SAGL_YEAST STANDARD: PRT; 650 AA.
 AC P20840;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Alpha-agglutinin precursor (AG-alpha-1).
 GN SAG1 OR AGA11 OR YJR004C OR J1418.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90014768; PubMed=2677666;
 RA Lipke P.N., Wojciechowski D., Kurjan J.;
 RT "AG alpha 1 is the structural gene for the Saccharomyces cerevisiae
 RT alpha-agglutinin, a cell surface glycoprotein involved in cell-cell
 RT interactions during mating";
 RL Mol. Cell. Biol. 9:3153-3165(1989).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX STRAIN=S288C;
 RC MEDLINE=90005993; PubMed=2676603;
 RA Hauser K., Tanner W.;
 RT "Purification of the inducible alpha-agglutinin of S. cerevisiae and
 RT molecular cloning of the gene";
 RL FEBS Lett. 255:290-294(1989).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX STRAIN=S288C / FY1679;
 RA de Haan M., Smits P.H.M., Grievell L.A.;
 RL Submitted (May-1995) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP PARTIAL SEQUENCE, CARBOHYDRATE-LINKAGE SITES, AND DISULFIDE BONDS.
 RX MEDLINE=96064684; PubMed=7592821;
 RA Chen M.-H., Shen Z.-W., Bobin S., Kahn P.C., Lipke P.N.;
 RT "Structure of Saccharomyces cerevisiae alpha-agglutinin. Evidence for
 RT a yeast cell wall protein with multiple immunoglobulin-like domains
 RT with atypical disulfides";
 RL J. Biol. Chem. 270:26168-26177(1995).
 CC -1- FUNCTION: CELL SURFACE GLYCOPROTEIN PROMOTING CELL-CELL CONTACT
 CC TO FACILITATE MATING. SACCHAROMYCES CEREVISIAE A AND ALPHA CELLS
 CC EXPRESS THE COMPLEMENTARY CELL SURFACE GLYCOPROTEINS A-AGGLUTININ
 CC AND ALPHA-, RESPECTIVELY, WHICH INTERACT WITH ONE ANOTHER TO
 CC PROMOTE CELLULAR AGGREGATION DURING MATING.
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
 CC (POSSIBLE).
 CC -1- INDUCTION: BY EXPOSITION TO PHEROMONE (A-FACTOR) SECRETED BY THE
 CC OPPOSITE MATING TYPE CELLS (TYPE A).
 CC -1- PTM: N-GLYCOSYLATED AND O-GLYCOSYLATED.

CC -1- SIMILARITY: TO C.ALBICANS ALST.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL contribution -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib.ch).
 CC -----
 DR EMBL: M28164; AAA34417.1; -
 DR EMBL: X16861; CAA34752.1; -
 DR EMBL: X87611; CAA60926.1; -
 DR EMBL: Z49504; CAA89526.1; -
 DR PIR: S22835; S22835.
 DR SGD: S0003764; SAG1.
 KW Glycoprotein; Cell adhesion; Signal; GPI-anchor; Repeat.
 FT SIGNAL 1 19
 FT CHAIN 20 650
 FT DOMAIN 278 350
 FT FT
 FT FT
 FT DOMAIN 339 423
 FT REPEAT 339 378
 FT REPEAT 384 423
 FT DISULFID 97 114
 FT DISULFID 202 300
 FT CARBOHYD 79 79
 FT CARBOHYD 109 109
 FT CARBOHYD 135 135
 FT CARBOHYD 148 148
 FT CARBOHYD 248 248
 FT CARBOHYD 282 282
 FT CARBOHYD 289 289
 FT CARBOHYD 299 299
 FT CARBOHYD 303 303
 FT CARBOHYD 306 306
 FT CARBOHYD 307 307
 FT CARBOHYD 308 308
 FT CARBOHYD 311 311
 FT CARBOHYD 314 314
 FT CARBOHYD 315 315
 FT CARBOHYD 316 316
 FT CARBOHYD 329 329
 FT CARBOHYD 331 331
 FT CARBOHYD 334 334
 FT CARBOHYD 335 335
 FT CARBOHYD 338 338
 FT CARBOHYD 339 339
 FT CARBOHYD 340 340
 FT CARBOHYD 341 341
 FT CARBOHYD 342 342
 FT CARBOHYD 345 345
 FT CARBOHYD 346 346
 FT CARBOHYD 349 349
 FT CARBOHYD 350 350
 FT CARBOHYD 364 364
 FT CARBOHYD 402 402
 FT CARBOHYD 460 460
 FT CARBOHYD 485 485
 FT CARBOHYD 501 501
 FT CARBOHYD 614 614
 FT CONFLICT 449 449
 FT CONFLICT 556 556
 FT CONFLICT 581 581
 SQ SEQUENCE 650 AA: 70339 MW: 88BF7A1C44C93C2B CRC64;
 Query Match Score 114; DB 1; Length 650;
 Best Local Similarity 20.9%; Pred. No. 0.24;
 Matches 98; Conservative 52; Mismatches 147; Indels 172; Gaps 23;
 OY 4 ILFIPLFFSSVL-----VFPA-----VSADKIPG-----DESTITNFGRRDNE--- 43
 DB 8 IIMFLSLAASAININDITFSNLEITPITANKOPDGGWTATPDSIASISIREGDEFTL 67

```

OY 44 SSP---KHNLNNHITR-----YSESHLYDRMFLC-----LSHNTLN 80
DB 68 SMHVVYIKRLKLNSSQRTISLADGTEAFKCYVSOQAAYLENTTFTCTAONDLSSTYPTD 127
OY 81 GACPTSENPSSSSVSGFTNTITTOFTKRSLLIKRELOIKGYKOLLFKSVNC-PSGLTLNSA 139
DB 128 GSITFSINFGDSSSYEYL-----ENAKFFKSGPMLVKLGONMSDVVNDPAFTENVE 182
OY 140 HFNCNKAAGASLYL--YIPAGELKNLPFGGIMDATLKLKRVKRRYSEYTGITINTIK 197
DB 183 HSGRSTGYGFSFESYHLGMYCPNGYF---LGG-----TEKI-----DVSNNNDL- 225
OY 198 LTFKGNIOI-----MLPFKSDARVDLNLPRGCGTYIGRN---SYDMCFYDGS--- 242
DB 226 --DCSSVOYVSSNDPNDMPPOSTINDTADY-----TGFSGNMLTTIDKLYDGEML 275
OY 243 -----YSTNSSSLE--IRPDNN--PKSDGKF 265
DB 276 WVALQSLPANVNTIDHLEFOYTCUDTIANITYATQFSTTRRETYOQGNLGLASAKSS 335
OY 266 YLRKINDTKEI---AYTL-----LLAGKSLPTNGTSLINLADAA 304
DB 336 FISTTTTDLTSIMTSAVSTGISTVETGNRTSEVISHVYTTSTKLSPATTSITIAOTS 395
OY 305 SLETFNN-----RTAVTMEPISVPVLCWPGRL 332
DB 396 IYSTDNITVGTIDHTTSEVISDETISTRETASTVAAPTSTGTWGAM 444

RESULT 3
GAS3_YEAST
AC 003655; STANDARD; PRT; 524 AA.
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, last sequence update)
DT 01-MAR-2002 (Rel. 41, last annotation update)
DE GAS3 protein precursor.
GN GAS3 OR YMR215W OR YMR261.09.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=528BC / AB972;
RA Dedman K., Brown D., Bowman S., Barrell B.G., Rajandream M.A.,
RA Walsh S.V.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP IDENTIFICATION.
RX MEDLINE=20529944; PubMed=11079560;
RA Pardo M., Ward M., Bains S., Molina M., Blackstock W., Gil C.,
RA Nombela C.;
RT "A proteomic approach for the study of Saccharomyces cerevisiae cell
RT wall biogenesis";
RL Electrophoresis 21:3396-3410(2000).
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC (Potential).
CC -1- SIMILARITY: BELONGS TO THE GAS1 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL 249809; CAAB9930.1;
CC COMPUYEAST-2DPAGE; 003655;
CC SCD: S0004828; GAS3.
CC Glycoprotein; Membrane; GPI-anchor; Signal.
KW

```

```

FT SIGNAL 1 21
FT CHAIN 22 524
FT PROPEP 2 524
FT DOMAIN 458 496
FT CARBOHYD 201 201
FT CARBOHYD 269 269
FT CARBOHYD 350 350
FT CARBOHYD 385 385
FT CARBOHYD 404 404
FT CARBOHYD 422 422
FT CARBOHYD 504 504
SQ SEQUENCE 524 AA; 56793 MW; 3A2BEDDBD3ED8690 CRC64;

Query Match 5.7%; Score 108; DB 1; Length 524;
Best Local Similarity 24.7%; Pred. No. 0.53;
Matches 69; Conservative 42; Mismatches 96; Indels 72; Gaps 18;

OY 119 GYKQL--LFSVNCPSGLTNSAHFNCNKNA---SGASLYLYIPAGELKNLPFGGIMDA 173
DB 264 GYDKLNSTFEDAVIP-----LIFSEYGCNKNTPTFDEYSEGL---GGLNVPFSGGL--- 313
OY 174 TLKLRVRRYSEYGYTITITIKLTDKNGIQIWLPOFKSD-ARVDLNLK---PTGGGT 228
DB 314 -----YETTEANNYGL---VKLDDSGSL-----TYKDDFVNLFSQLNKVSLEPTKES 359
OY 229 YIGRNSVDMC-----FYGCTNSSSLE-----IRPDNNKSDGK----- 264
DB 360 EISDSIYKCDNSAITNIYSGFTNFTLPSQPAEIANMIEYGVNGTNT-GKILTDYAVP 418
OY 265 ----FYLRKINDPTKEIAYTILSLLAGKSLPTNGTSLINLADLSLEPTNNNR--TTAVTNP 319
DB 419 TTFYTTIKNNKDDT--ISAFTISYDKA-NSLNELDVYATTTAKSASTSQSSRSRLTSTSP 475
OY 320 EISVPVLCWPGRLQLDKAVENPEAGYMGNI-NVTFTPS 357
DB 476 SSSWGSSSSSTG---SSSASSSSKSKGVNIVNVFSFGS 510

RESULT 4
YDEO_ECOLI
AC P77588; STANDARD; PRT; 304 AA.
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, last sequence update)
DT 16-OCT-2001 (Rel. 40, last annotation update)
DE Hypothetical fimbrial-like protein ydeQ precursor.
GN YDEO OR B1502.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blather F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:12453-12474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97251357; PubMed=9097039;
RA Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Miki T., Mizoduchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nashimoto H., Nishio Y., Oshima T., Saito N.,
RA Sempel G., Seki Y., Sivasubraman S., Tagami H., Takeda J.,
RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome

```

RT corresponding to the 28.0-40.1 min region on the linkage map.
 RL DNA Res. 3:363-377(1996).
 CC -1- SIMILARITY: TO TYPE-1 FIMBRIAL SUBUNITS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AEO00247; AAC74575.1; .
 DR EMBL: D90792; BAA15175.1; .
 DR EMBL: D90793; BAA15183.1; .
 DR EcoGene: EG13799; ydeQ.
 DR InterPro: IPR000259; Fimbril.
 DR Pfam: PF000419; Fimbril; 1.
 DR Hypothetical protein; Fimbril; Signal; Complete proteome.
 KM SIGNAL
 FT CHAIN 1 26
 FT SIGNAL
 FT CHAIN 27 304
 FT SIGNAL
 SQ SEQUENCE 304 AA; 32069 MW; 8153C86E3087D99A CRC64;

Query Match 5.7%; Score 107; DB 1; Length 304;
 Best Local Similarity 21.0%; Pred. No. 0.32;
 Matches 75; Conservative 51; Mismatches 125; Indels 106; Gaps 19;

QY 3 KILF-IPTLFSSVLEFAVSADKIPGDESI---TNIFGRPRNESPKNHT---LNH 54
 DB 8 KVFGLYLLMAGKVFNFSCNV---GSSIGAGTTSVYVNL-DPVIOQNLVVDLSQH 63
 QY 55 ITAVSESHLYDRMTFLCLSSH-NTINGACPTSENPSSSVSGETNTITLOFTEKRLIKR 113
 DB 64 ISCMNDYGMWD-----TDHNLVYG-----SAFAG----- 89
 QY 114 ELQIKGK-QLKRSVNCPCGLTNSAHFNCNKAASGASLYIYPAGELKNLPFGIWD 172
 DB 90 -SLQSGKSLYNNVNTYPPPLTNTNVDIGDKTPMPLKLYI-----TPVGAAG 140
 QY 173 ATLK-----LRVRRYSEFYGT-----YTINIRK-----LTDKNIDIMLPQK 212
 DB 141 VVKAEEVARIIMYKATIGSGNPRNFTWNITSNNNVMPGCGCYVDSKNVVDLPDP 200
 QY 213 SDARVDLNLPTGGTYIGRNSVDMCFYDGYSTNSSLEIRFODNNPKSGKFLRKIND 272
 DB 201 GSAEIPL-----GYVCS-SEQKLSFYLSCATTDSRQV-FANTAP-----D 239
 QY 273 DREIAYTSLSLAGK-----SLTPNGTSLNIAADAASLETNNRITAVTMEI 321
 DB 240 ATKASGVGLMNRKILATGENVSLGTVMKSKVPLGLSATYQGTGNKVSAGTVQSV 296

RESULT 5

ALYS_ENTR 5
 ID Alys_ENTR STANDARD: PRT: 671 AA.
 AC P37710;
 DT 01-OCT-1994 (Rel. 30; Created)
 DT 01-OCT-1994 (Rel. 30; Last sequence update)
 DT 16-OCT-2001 (Rel. 40; Last annotation update)
 DE Autolysin precursor (EC 3.2.1.-) (peptidoglycan hydrolase)
 DE (Beta-1,4-galactosidase).
 OS Enterococcus faecalis (Streptococcus faecalis).
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;
 OC Enterococcus.
 OX NCBI_TaxID=1351;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91358349; PubMed=1679432;
 RA Bellevue C., Potvin C., Trudel J., Asselin A., Bellemare G.;
 RT "Cloning, sequencing, and expression in Escherichia coli of a
 RT Streptococcus faecalis autolysin."
 RL J. Bacteriol. 173:5619-5623(1991).

CC -1- FUNCTION: HYDROLYSES THE CELL WALL OF E. FAECALIS AND
 CC M. LYSODEIKTUS. MAY PLAY AN IMPORTANT ROLE IN CELL WALL GROWTH
 CC AND CELL SEPARATION.
 CC -1- SUBCELLULAR LOCATION: Secreted (Probable).
 CC -1- DOMAIN: LYSM REPEATS ARE THOUGHT TO BE INVOLVED IN PEPTIDOGLYCAN
 CC BINDING.
 CC -1- SIMILARITY: BELONGS TO FAMILY 73 OF GLYCOSYL HYDROLASES.
 CC -1- SIMILARITY: CONTAINS 5 LYSM REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M58002; AAB67325.1; .
 DR PIR: A38109; A38109.
 DR InterPro: IPR002901; Amidase_4.
 DR InterPro: IPR002482; LysM_4.
 DR Pfam: PF01476; LysM; 5.
 DR Pfam: PF01476; Amidase_4; 1.
 DR SMART: SM00047; LysM; 5.
 DR SMART: SM00257; LysM; 5.
 DR SMART: SM00257; LysM; 5.
 DR Hydrolase; Glycosidase; Bacteriolytic enzyme; Cell wall;
 KM Cell division; Septation; Repeat; Signal.
 FT SIGNAL 1 53
 FT CHAIN 54 671
 FT SIGNAL
 FT CHAIN 63 405
 FT REPEAT 431 473
 FT REPEAT 499 541
 FT REPEAT 567 609
 FT REPEAT 629 671
 FT REPEAT 671 671
 SQ SEQUENCE 671 AA; 70442 MW; 3458268C6C1A433 CRC64;

Query Match 5.6%; Score 106; DB 1; Length 671;
 Best Local Similarity 23.6%; Pred. No. 1;
 Matches 73; Conservative 39; Mismatches 103; Indels 94; Gaps 17;

QY 41 RNESPKNHT-LNNHTAVSESHLYDRMTFLCLSHNTLNGACPTSENPSSSV----- 94
 DB 309 RYATDPSYNAKLNNTVITAY--NLQYDTPSSGNGTGGYVNPETGSNNQSGNTYYTVK 366
 QY 95 SGEF--NITLQF-----TEKSL--IKRELQIKYKOLLKRSVNCPCGLTNSAHFNCN 146
 DB 367 SGTLLKKAIAQYGVSVANLRSMNGISGDLFVQKLIYKGA--SGNTGGSGNGSGNN 423
 QY 147 AASGASLYIYPAGELKNLPFGIWDATLKRVRSEFYGTYYTINIRKLTDKNIOI 206
 DB 424 -QSGTNTYYTVKSGDTLN-----KIAAQY-----VIV-----ANLNS 455
 QY 207 WLPOKSDARVDNL-----RPTGGGTIGRNSVDMCFYDGYSTNSSLEIRFODNN 258
 DB 456 W-----NGISGDLFVQKLIYKKGTSNMT-----GGSSNGS-----NNN 491
 QY 259 PKSDGKFLYKINDDTKEIA--YTLN-----LLAGKSLTPNGTSLIAIA 303
 DB 492 QSGTNTYYTVKSGDTLNKKAIAQYGVSVANLRSMNGISGDLFVQKLIYKKGTSN 551
 QY 304 ASLETNNNR 312
 DB 552 SNGGSNNQ 560

RESULT 6

IGA4_HAEIN
 ID IGA4_HAEIN STANDARD: PRT: 1849 AA.
 AC P45386;
 DT 01-NOV-1995 (Rel. 32; Created)
 DT 01-NOV-1995 (Rel. 32; Last sequence update)
 DT 01-NOV-1995 (Rel. 32; Last annotation update)

[illegible]

OY 229 YIGNSVDMCYDGYSTNSSLEIRFQDNNKSGKFEYLRKINDTKEIAYT-----LST 283
 Db 259 YTNRTGTLMKY--GFAT-----IFFENGTGKSTKI-----EGIATGYGVSA 300
 OY 284 LLAGKSLTPNG-----TSINTADASLETNNKRIATVMPET--SVPULCWGRIQL 334
 Db 301 TKESPFLNASSGKYEIYESANVSNKAS--SYEPNLTHVTIMAVNGNPILDEPNTL 357

RESULT 8
 SP21_YEAST STANDARD: PRT: 758 AA.

AC P35209;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE SPT21 protein.
 GN SPT21 OR YMR179W OR YMR010.09.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94186069; PubMed=8138180;
 RA Natsoulis G., Winston F., Boeke J.D.;
 RT "The SPT10 and SPT21 genes of Saccharomyces cerevisiae.";
 RL Genetics 136:93-105(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / AB972;
 RA Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
 RL Submitted (JUN-1995) to the EMBL/GenBank/DDJ databases.
 RT -I- FUNCTION: REQUIRED FOR NORMAL TRANSCRIPTION AT A NUMBER OF LOCI IN
 CC YEAST.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: L24436; AAA3078.1; -
 DR EMBL: 249808; CAA89912.1; -
 DR PIR: S47866; S47866.
 DR SGD: S0004791; SPT21.
 FT DOMAIN 127 144 ASF/GLU-RICH (ACIDIC).
 FT DOMAIN 672 682 ASF/GLU-RICH (ACIDIC).
 SQ SEQUENCE 758 AA; 84697 MW; 7DB3FCFE96705 CRC64;

Query Match 5.2%; Score 98; DB 1; Length 758;
 Best Local Similarity 21.9%; Pred. No. 5;
 Matches 61; Conservative 34; Mismatches 95; Indels 88; Gaps 12;

OY 42 NESSPKHNITNNHTAYSESHLYDKMTF-----LCSSHNLTNGACPSENSSSVS 95
 Db 166 NISNKKGRVNNNOI-----PETLEVKLRFTKYITNLTSNNNTNSRISCLQWPSL-- 218
 OY 96 GETNITLOETEKRSLLKRELQIKGYKQLFKSYNCPSGLTNSAHFNCKNNAASGLYL 155
 Db 219 --PSATLPLFPKQSOLFETKQIK-----NSRNARTITIIIN-----NTNSGT-- 257
 OY 156 YIPAGELKNLPGGIWDATLKLKLVKRRYSFTYTYTINITIKLTKGNQIQIWLDPQEKSDA 215
 Db 258 --VGRROTNPM-----PAPKAVRTO-----SLPIW----- 280
 OY 216 RVDLNRPTGGGTYIGRNSDMCFY-----DGYSTNSSLEIRFQDNNKSGKFEYLRK 269
 Db 281 ----NLKPNIANGPPRNSIAHKIYIADRKTEANQONHONITAYEINTLQNDNTIORTK 336

OY 270 INDD-TKEIAYTSLLAGKSLTPNGT-----SLNI 300
 Db 337 IDDSVSKRPFDMLNKRKRSKTKVSPGIATIKKPRASINI 374

RESULT 9
 N133_YEAST STANDARD: PRT: 1157 AA.

AC P36161;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Nucleoporin NUP133 (Nuclear pore protein NUP133).
 GN NUP133 OR YKR082W OR YKR402.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=JUX1R;
 RX MEDLINE=95112817; PubMed=7813444;
 RA Doye V., Wept R., Hurt E.C.;
 RT "A novel nuclear pore protein Nup133p with distinct roles in poly(A)+
 RT RNA transport and nuclear pore distribution.";
 RL EMBO J. 13:6062-6075(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94262327; PubMed=8203164;
 RA Garcia-Cantalejo J., Baladron V., Esteban P.F., Santos M.A., Bou G.,
 RA Remacha M.A., Revuelta J.L., Ballesta J.P.G., Jimenez A., del Rey F.;
 RT "The complete sequence of an 18,002 bp segment of Saccharomyces
 RT cerevisiae chromosome XI contains the HBS1, MRP-120 and PRP16 genes,
 RT and six new open reading frames.";
 RL Yeast 10:231-245(1994).
 RN [3]
 RP FUNCTION: INVOLVED IN POLY(A)+ RNA TRANSPORT AND NUCLEAR PORE
 CC DISTRIBUTION.

CC -----
 CC -I- SUBCELLULAR LOCATION: Nuclear pore complex.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X80066; CAA56372.1; -
 DR EMBL: 227116; CAA81633.1; -
 DR EMBL: 228307; CAA82161.1; -
 DR PIR: S38160; S38160.
 DR PIR: S39123; S39123.
 DR SGD: S0001790; NUP133.
 DR Nucleic acid: Transport; Transmembrane.
 KW Nuclear protein; POTENTIAL.
 FT TRANSMEM 217 233 TO NUP120 (AA 434-763).
 FT SIMILAR 98 420
 SQ SEQUENCE 1157 AA; 133319 MW; C8BDBB7D709C5C08 CRC64;

Query Match 5.2%; Score 98; DB 1; Length 1157;
 Best Local Similarity 22.1%; Pred. No. 8.5;
 Matches 75; Conservative 43; Mismatches 127; Indels 94; Gaps 21;

OY 48 HNLTNNHTAY-----SESHLYDKMTF-----LCSSHNLTNGACPSEN 89
 Db 96 YGLVNHKKVYIWNHSTOKDPTIYIVPFRSDNDDEIAVAPRCILTFPATMDESPALNP 155
 OY 90 SSSVSAGE-----TNITLOETEKRSLLKRELQIK-----GYKQLRK 126
 Db 156 NDDDETGGLIITIGSKAIYIEDINSLNNLNFKLSERFS--HELELPINSSGGEKCDLM- 211
 OY 127 SYNC-PSGLTNS--AHFNCKNNAASGLYLYIPAGELKNLPGG-GIWDATIKLR-- 178

Matches	65;	Conservative	43;	Mismatches	131;	Indels	85;	Gaps	12;
---------	-----	--------------	-----	------------	------	--------	-----	------	-----

```

QY      68  MTEFLCLSSHTNTINGACPPSENSSSSVSGETNITLQTEKRSKLKRELQIKGYOLFKS 127
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      8  LTSVSMSYIGIALGALPWPASQ-DRSVANNKLIIFS---IDTVQIYTVDRFDVG-----NS 57
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      128 VNCPSGGLNLNSHFNCNNKMAASGASLYLYIRAGELKNLPFGCINWATLTKRKRYSELY 187
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      58 ANNPPTGAFFSSDHSN-----Lkly-----FGDDMGQ-----ITNKINDGY 92
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      188 GMYTINITIKLDDKNINIOIWLPOFSDARVDLNLPTGGGYIG----- 231
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      93 -----LTGMGITALMISQPYENITAVINYSGVANNATVHGVPWPRDFKTKNAAPRSF 142
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      232 -----RNSVDMCFYDGYSTNSSLEIRFODNPNPKSDGKFLYLRKINDPTKEIA 278
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      143 TDFESMLIAAHSHNKIKYVWDFAPNHTNPASTSDPSFAEGALYNNGTLLGKYSNDTAGLE 202
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      279 YTLISLLAGKSLTPP-NGTSLNIADPALETFWNNITAVTMEISVPVLCWPGRLQIDAK 337
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      203 HH-----NGTDFSTESGIYKNLYDLADIQNNNNITDLSYKESIQIWLMLGVGDGIRPDV 258
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      338 VENPEAGQ--YWGNI-----NWTF 355
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      259 KHPQGMOKRSYVSSLYSSANPVFT 282
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 11
REF1M_YEAST
AC      P30775;
DT      01-JUL-1993 (Rel. 26, Created)
DT      01-JUL-1993 (Rel. 26, Last sequence update)
DT      01-MAR-2002 (Rel. 41, Last annotation update)
DE      Peptide chain release factor 1, mitochondrial precursor (MRF-1).
GN      MRF1 OR Yel143C.
OS      Saccharomyces cerevisiae (Baker's yeast).
OC      Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC      Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX      NCBI_TaxID=4932;
      [1]
RN      SEQUENCE FROM N.A.
RP      MEDLINE=93117110; PubMed=1475194;
RA      Pel H.J., Maat M.J., Rep M., GriVe11 L.A.;
RT      "The yeast nuclear gene MRF1 encodes a mitochondrial peptide chain
      release factor and cures several mitochondrial RNA splicing
      defects.";
RL      Nucleic Acids Res. 20:6339-6346(1992).
      [2]
RN      SEQUENCE FROM N.A.
RP      STRAIN=S288C / FY1769;
RA      MEDLINE=97197983; PubMed=9046099;
RT      Voelt M., Defoor E., Verhasselt P., Riles L., Robben J., Voickaert G.;
RT      "The sequence of a nearly unclonable 22.8 kb segment on the left arm
      chromosome VII from Saccharomyces cerevisiae reveals APO2, RPL3A,
      TPI1, MRF1 genes and six new open reading frames.";
RL      Yeast 13:177-182(1997).
-1- FUNCTION: MITOCHONDRIAL PEPTIDE CHAIN RELEASE FACTOR THAT DIRECTS
      TERMINATION OF TRANSLATION IN RESPONSE TO THE PEPTIDE CHAIN
      TERMINATION CODONS UAA AND UAG.
-1- SUBCELLULAR LOCATION: Mitochondrial.
-1- SIMILARITY: BELONGS TO THE PROKARYOTIC AND MITOCHONDRIAL RELEASE
      FACTORS FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
EMBL; X60381.CAA42932.1; -.

```


FT REPEAT 2028 2047 2-27.
 FT REPEAT 2051 2070 2-28.
 FT REPEAT 2071 2090 2-29.
 FT REPEAT 2093 2112 2-30.
 FT REPEAT 2120 2139 2-31.
 SQ SEQUENCE 2334 AA; 258329 MW; B75138CCD278BA3 CRC64;

Query Match 5.0%; Score 95; DB 1; Length 2334;
 Best Local Similarity 22.1%; Pred. No. 35; Mismatches 106; Indels 102; Gaps 20;

Matches 71; Conservative 42; Mismatches 106; Indels 102; Gaps 20;

CC 29 DESITNIFGPRDRNESSPHNIIILN-----HTTA-----YSESHTY-----DEMTFLCLIS 74
 DB 260 DSKLDLSEVERSD-KVSYKLEKNEGYLHLITADENMLKDERYKYPISIDSTLSIS 318
 QY 75 SHNTLNGACPTSENPSSSSVGETNITLQFTKRSILIKRELQIKGYKOLLFKSVNCPSGI 134
 DB 319 SDFVMSAVPTT-NYSASSQKMDAN-----LKAAY----- 346
 QY 135 TLNSAFNCKNNAASGASLYIPIAGELKNLPRGIMDATIKLRVRRYSETYGTITIN 194
 DB 347 VLKGTGY-----DKTGTN-YAFMKFNMLKPIQMTVTKATLKTYVAHSY---YGT----- 393
 QY 195 TIKLTKGNIOIWLPOFKS---DARYDLNLPRPGGTYIGRNSVDMCFYDGY-----S 244
 DB 394 --KATG-----LMLDTVNSNYDNAKYMTWKPKASKN--IGKADYHKGQMSYDVTAAVKS 444
 QY 245 TNSSSLEIFRDNPNKSDGKFLYLRKI-----NDDTKELAVTLLSLAGKSLTPT----- 293
 DB 445 WNSGANGYGR-LHTNGCKEYWKRLISSANSANKPIEYTYIIP-----KGMTPIIKAY 498
 QY 294 -NCTSLNIDAAASLETNMNR 313
 DB 499 HNGDSTGYPD-----ISMKV 514

RESULT 13
 N100_YEAST
 ID N100_YEAST STANDARD; PRT; 959 AA.
 AC 002629;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Nucleoprotein NUP100/NSP100 (Nuclear pore protein NUP100/NSP100).
 GN NUP100 OR NSP100 OR YKL068W OR YKL336.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93054906; PubMed=1385442;
 RA Wente S.R.; Rout M.P.; Blobel G.;
 RT "A new family of yeast nuclear pore complex proteins.";
 RL J. Cell Biol. 119:705-723(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C;
 RX MEDLINE=94378724; PubMed=80091863;
 RA Rasmussen S.W.;
 RT "Sequence of a 20.7 kb region of yeast chromosome XI includes the
 NUP100 gene, an open reading frame (ORF) possibly representing a
 nucleoside diphosphate kinase gene, tRNAs for His, Val and Trp in
 addition to seven ORFs with weak or no significant similarity to
 known proteins.";
 RL yeast 10:569-574(1994).
 CC -!- FUNCTION: ESSENTIAL COMPONENT OF NUCLEAR PORE COMPLEX.
 CC NUCLEOPORINS MAY BE INVOLVED IN BOTH BINDING AND TRANSLLOCATION OF
 CC THE PROTEINS DURING NUCLEOCYTOPLASMIC TRANSPORT.
 CC -!- SUBCELLULAR LOCATION: Nuclear pore complex.
 CC -!- DOMAIN: CONTAINS G-L-F-G REPEATS.
 CC -!- SIMILARITY: BELONGS TO THE GLFG FAMILY OF NUCLEOPORINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

CC EMBL: Z15035; CAA78753.1; -
 DR EMBL: X75780; CAA53406.1; -
 DR EMBL: Z28068; CAA81905.1; -
 DR PIR: B44402; B44402.
 DR PIR: S39173; S39173.
 DR PIR: S44518; S44518.
 DR SGD: S0001551; NUP100.
 DR InterPro: IPR004325; Nucleoporin_Fg.
 DR Pfam: PF03093; Nucleoporin_Fg; 24.
 KW Nuclear protein; Transport; Repeat.
 FT DOMAIN 33 571
 FT SEQUENCE 959 AA; 99988 MW; D3985F9901BBA51 CRC64;

Query Match 5.0%; Score 94.5; DB 1; Length 959;
 Best Local Similarity 19.5%; Pred. No. 12; Mismatches 157; Indels 115; Gaps 17;
 Matches 80; Conservative 58; Mismatches 157; Indels 115; Gaps 17;

QY 31 SITNIFGPRDRNESS-----PKHNL-----NNHITVSESHTYDDEMTFLCLSSHTLNG 81
 DB 284 SNGSLFGQNSMNSSTIGVGQNNNOINOINNNSNLFKANTSNASGSLFGQNNQOG 343
 QY 82 ACPTSENPSSSVSG-----ETNITLQFTKRSILIKRELQIKGY- 121
 DB 344 SGLEFGQNSQTSFGSSGLFGQNNOKOPNTFTQSNIGTL-FQNNNOOQOSTGLGAPAGT 402
 QY 122 -----QLFKSVNCPGSLTLNSAHFNCKNNAASGASLYIPIAGELKNLPRG 168
 DB 403 TGSLEFGNSSTQPSLFGTNTVPTS-----NTQSQGNSLP--GATKLTMPG 449
 QY 169 GIMDATIKLRVRRYSETYGTITINITIKLTKGNIOIWLPOFKS----- 213
 DB 450 GNFTA-----NQSQSGNSLFGTRKASTGSLFGNNTASTYVPSYNGLFQNNANSTSTTN 505
 QY 214 ---DARYDLNLPRT-GGGTY-----IGRN-----SDMCTYDGTSTSSSL--- 250
 DB 506 GLEFARPDQSKPALGGLFGNSNSNSTIGQKPVFGTQNTGLFGATGTNSSAVGSY 565
 QY 251 -EIRPDNNPKSDGKFLYLRKINDTKEIAYTJSLLAGKSLRPTNGTSLNIDAAASLETN 309
 DB 566 GLEFGQNNNTLVNGTQNVPPVNTTQN-----ALGTTAVP---SLQAQAV---TN 610
 QY 310 WNRITATVMP-EISVPLVCPGRQLQDAKVENPEAQYMGINIVTTPSS 358
 DB 611 EQLFKISISIPNITNPKATTSKVNADMRNSSLTAVRLAPKPLRAPS 660

RESULT 14
 LEU2_BUCRP
 ID LEU2_BUCRP STANDARD; PRT; 471 AA.
 AC P48573;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE 3-isopropylmalate dehydratase large subunit (EC 4.2.1.33)
 DE (isopropylmalate isomerase) (Alpha-IPM isomerase) (IPM1).
 GN LEU2.
 OS Buchnera aphidicola (subsp. Rhopalosiphum padi).
 OG Plasmid PRPE.
 OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
 OX NCBI_TaxID=98793;
 RN [1]
 RP SEQUENCE FROM N.A.

RA MEDLINE=95333198; PubMed=7608990;
RA Briscoe A.M., Martinez-Torres D., Moya A., Iatorre A.;
RT "Discovery and molecular characterization of a plasmid localized in
RT Buchnera sp. bacterial endosymbiont of the aphid Rhopalosiphum
RT pad...";
RL J. Mol. Evol. 41:67-73(1995).
CC -1- CATALYTIC ACTIVITY: 3-ISOPROPYLMALATE = 2-ISOPROPYLMALATE +
CC H(2)O (ALSO CATALYZES 2-ISOPROPYLMALATE + H(2)O = 3-HYDROXY-
CC 4-METHYL-3-CARBOXPENTANONE).
CC -1- PATHWAY: SECOND STEP IN LEUCINE BIOSYNTHESIS.
CC -1- SUBUNIT: CONSISTS OF TWO DIFFERENT SUBUNITS: LEUC AND LEUD.
CC -1- SIMILARITY: BELONGS TO THE ACONITASE/IPM ISOMERASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X71612; CAA50617.1; -
DR InterPro: IPR001030; Aconitase.
DR Pfam: PF00330; aconitase; 1.
DR PRINTS: PR00415; ACONITASE.
DR PRODOM: PD000511; Aconitase; 1.
DR PROSITE: PS00450; ACONITASE_1; 1.
DR PROSITE: PS01244; ACONITASE_2; 1.
DR Leucine biosynthesis; Lyase; Iron-sulfur; 4Fe-4S; Plasmid.
KW METAL 347 347
FT METAL 409 409 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
FT METAL 412 412 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
SQ SEQUENCE 471 AA; 52596 MW; E509ABC23AFCE297 CRC64;

Query Match 5.0%; Score 94; DB 1; Length 471;
Best Local Similarity 22.3%; Pred. No. 5.5;
Matches 71; Conservative 46; Mismatches 98; Indels 104; Gaps 18;

OY 44 SSKPNNILNHTAYSEHTLYDRMTFLCLSSHNLTNGAC-PTSEPPSSSSVSGENITL 102
DB 101 NNPKGGIV-HVIGPEQGTLL-PGMTIVCGDSHTGAFGALSFEGIGTSEV-EHVLVT 155
OY 103 QPTEKSLIKRELQIKGYKO-----LTF-----KSNCPSS 132
DB 156 QTLKQORLKNMKIOTIKGKKNFTAKDILFTIIGLGTSGSGGYVEFGVDYIKKNMEE 215
OY 133 GLTNSAHFNCKNNAASGA-----SLYLYIPAGELKNLPFGGIMD-----ATLKL 177
DB 216 RMTV-----CNMAIEMGAKSALIPDETTFYLGKRRYS--PGKFWEEAVKYMKTLLT 267
OY 178 RKRRKRYSE-----TYGT---YTINITIKLTP-----KGIQIWLPOFKSDANV 217
DB 268 RSKAFPFDEKFTEDVSNLSQITWGTSPDOVLISINEKIPDFNFKSVKKNLRSACDY-- 325
OY 218 DNLAPPGGGTYIGRNSVDMCFYDGYSTSSSLEIRFQONNPKSDOKFKLRKINDTKEI 277
DB 326 -MNLKP---GSLTLDIKIDKFTI-GSCTRISRIEDLSRAKIILKN-----KTSIKIKAI 374
OY 278 AYTLSLLAGKSLTPNGT 296
DB 375 -----LLPESGT 381

RESULT 15
THBG_SHEEP
ID THBG_SHEEP STANDARD: PRT; 412 AA.
AC P50450;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Thymoxine-binding globulin precursor (T4-binding globulin).
GN SERPIN7 OR TBG.

OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxId=9940;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=94192879; PubMed=8143919;
RA "Sheep Thymoxine-binding globulin: cDNA sequence and expression.";
RL Mol. Cell. Endocrinol. 98:91-97(1993).
CC -1- FUNCTION: MAJOR THYROID HORMONE TRANSPORT PROTEIN IN SERUM.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- TISSUE SPECIFICITY: SYNTHESIZED IN THE LIVER; FOUND IN PLASMA.
CC -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X69795; CAA49450.1; -
DR HSSP: P01011; ZACH.
DR InterPro: IPR000215; Serpin.
DR Pfam: PF00079; serpin; 1.
DR SMART: SM00093; SERPIN; 1.
DR PROSITE: PS00284; SERPIN; 1.
KW Serpin; Glycoprotein; Plasma; Signal.
FT SIGNAL 1 16
FT CHAIN 17 412
FT CARBOHYD 35 35
FT CARBOHYD 98 98
FT CARBOHYD 164 164
FT CARBOHYD 252 252
SQ SEQUENCE 412 AA; 45948 MW; E420D3530D1A04D9 CRC64;

Query Match 5.0%; Score 93.5; DB 1; Length 412;
Best Local Similarity 18.4%; Pred. No. 5.1;
Matches 56; Conservative 54; Mismatches 102; Indels 93; Gaps 14;

OY 33 THIEGPRONNESSPKNNILNHTAYSES-----HTLYDRMFLCLSSHNLTNG 81
DB 155 TEVFSTDESNVSAAOEI-NSHVERQTKGIVGLIQDLKPNITIVLVNLTCKKA-QWAMP 212
OY 82 ACPTSENPSSSVSGETNTTQFTKRLIKRELQIKGYKQLPFSVNCPSGLTNSAHF 141
DB 213 FPPSKTEEGSSFLVDKTT-TVO-----VPMHGVDOYIHLVDLENC-----TVLQM 258
OY 142 NCKNNAASGASLYLYIPAGELKNLPFGGIMDATLRLVKRRYSERYGYTINITIKLTDK 201
DB 259 DYSKNA---LALFVLPKEGMEGV-----EGAMSSKIILKWN-----RLQK 297
OY 202 GNIOIWLPOFKSDANVDNLRFPG-
DB 298 GWNVLEVPKFSISATYD-----GGILLKMGIDAPADNADFSGLTKDNGLKVSNAVAKA 352
OY 238 CFYDGYSTSSSLEIRFQONNPKS-----DGKFLYLRKINDTKEIAYTSLILLAG 288
DB 353 MYEIGKGTALPEYARFL-NQPETILLHPTIOFDSFLLILKKNTR-----SLFPLCK 405
OY 289 SLRPT 293
DB 406 VVDPT 410

Search completed: July 1, 2002, 11:25:52
Job time: 128 sec

This Page Blank (uspto)

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 1, 2002, 11:23:04 ; Search time 28.51 Seconds
(without alignments)
2190.501 Million cell updates/sec

Title: US-09-839-894-10
Perfect score: 1886
Sequence: 1 MNKILFTLFFSSVLTFTFA.....EAGQYMGNTVTFPPSSQTL 361

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SPTREMBL_19:*

- 1: sp.archaea:*
- 2: sp.bacteria:*
- 3: sp.fungi:*
- 4: sp.human:*
- 5: sp.invertebrate:*
- 6: sp.mammal:*
- 7: sp.mhc:*
- 8: sp.organelle:*
- 9: sp.phage:*
- 10: sp.plant:*
- 11: sp.rodent:*
- 12: sp.virus:*
- 13: sp.vertibrate:*
- 14: sp.unclassified:*
- 15: sp.virus:*
- 16: sp.bacteriap:*
- 17: sp.archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1886	100.0	361	2	093G67
2	949.5	50.3	363	2	047J15
3	930.5	49.3	364	2	047J19
4	275.5	14.6	359	2	09XDS1
5	115.5	6.1	847	2	P71132
6	110	5.8	846	2	P71133
7	109	5.8	2349	2	P94750
8	109	5.8	2383	16	P76347
9	108	5.7	872	2	09RUA0
10	108	5.7	2529	16	025579
11	105.5	5.6	430	3	096VN2
12	104.5	5.5	1441	16	09CFL1
13	104	5.5	803	10	09AUS5
14	103.5	5.5	691	16	09RZS7
15	103.5	5.5	1269	17	097Z06
16	103.5	5.5	2399	16	09ZKS9

17	102.5	5.4	1736	5	095PH7	095PH7 dictyosteli
18	102.5	5.4	2747	2	091B00	091B00 aeromonas s
19	102	5.4	692	17	097YM6	097YM6 sulfolobus
20	100	5.3	1014	2	046962	046962 escherichia
21	100	5.3	1752	10	09LX60	091X60 arabidopsis
22	99.5	5.3	1300	16	09CK27	09CK27 pasteurella
23	99.5	5.3	5627	16	091120	091120 pseudomonas
24	99	5.2	456	2	048508	048508 lactobacilli
25	99	5.2	802	16	092C16	092C16 listeria in
26	99	5.2	839	2	P77792	P77792 chlamydia p
27	99	5.2	856	16	09PHX4	09PHX4 campylobact
28	98.5	5.2	4919	2	09ZHL0	09ZHL0 haemophilus
29	98	5.2	587	16	092E42	092E42 listeria in
30	98	5.2	1014	2	047106	047106 escherichia
31	97.5	5.2	855	15	066861	066861 feline immu
32	97.5	5.2	1753	17	026812	026812 methanother
33	97	5.1	334	2	0930W7	0930W7 bacteroides
34	97	5.1	454	10	081310	081310 arabidopsis
35	97	5.1	803	2	093430	093430 penicillium
36	96.5	5.1	2364	2	046342	046342 clostridium
37	96	5.1	406	2	09FOR4	09FOR4 mycoplasma
38	96	5.1	430	10	0932N5	0932N5 arabidopsis
39	96	5.1	992	8	021375	021375 pichia kluy
40	95.5	5.1	895	10	080973	080973 arabidopsis
41	95.5	5.1	2081	17	09HJ37	09HJ37 thermoplasma
42	95	5.0	198	2	09K541	09K541 mycobacteri
43	95	5.0	561	16	09K6M4	09K6M4 bacillus ha
44	95	5.0	867	2	093G68	093G68 escherichia
45	94.5	5.0	200	10	09SB15	09SB15 lagopyrum c

ALIGNMENTS

RESULT 1
ID 093G67 PRELIMINARY; PRF; 361 AA.
AC 093G67;
DT 01-DEC-2001 (TREMblrel. 19, Created)
DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE MINOR PILIN PROTEIN CSAE.
GN CSAE.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EL1881A;
RA Althoum Z.D., Levine M.M., Galen J.E., Barry E.M.;
RT "Isolation and characterization of EPEC CSA fimbriae encoding genes,
RT and their expression in Shigella flexneri 2a guaba strain CVD 1204.";
RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF296132; AAK97137.1;
SQ SEQUENCE 361 AA; 40102 MW; 2E1E74ABDB0EB353 CRC64;

Query Match	100.0%;	Score 1886;	DB 2;	Length 361;
Best Local Similarity	100.0%;	Pred. No. 1.1e-138;		
Matches 361;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
OY 1	MNKLFTLFFSSVLTFTFAVSADKIPGDEISITNFGPRDRNESSPKHNLINHTAYSE 60			
Db 1	MNKLFTLFFSSVLTFTFANASADKIPGDEISITNFGPRDRNESSPKHNLINHTAYSE 60			
OY 61	SHLYDRMTFICLSSAHNTLACPTSENPSSSVSGETNITLQFTKRSILKRELQIKGY 120			
Db 61	SHLYDRMTFICLSSAHNTLACPTSENPSSSVSGETNITLQFTKRSILKRELQIKGY 120			
OY 121	KOLFESVNCPSGLTNSAHFNCKNNAASGASLYLTPAGELKMLPGGTMDATLKLRYK 180			
Db 121	KOLFESVNCPSGLTNSAHFNCKNNAASGASLYLTPAGELKMLPGGTMDATLKLRYK 180			

```

OY 181 RRYSEYGYTINTITIKLTDKGNIOIWLPOFKSDARVDLNRPTGGTYIGRNSVDMCFY 240
    |||:|||||:|
DB 181 RRYSEYGYTINTITIKLTDKGNIOIWLPOFKSDARVDLNRPTGGTYIGRNSVDMCFY 240
OY 241 DGYSTNSSLEIRFODNNPKSDGKFYLRKINDDTKEIAYTLLSLAKSLTPPINGTSLNT 300
    |||:|||||:|
DB 241 DGYSTNSSLEIRFODNNPKSDGKFYLRKINDDTKEIAYTLLSLAKSLTPPINGTSLNT 300
OY 301 ADAASLETNNRITAYTAVTMEISVAVLCPGRLOLDKAVENPEAGQYMGINVTFTSSQT 360
    |||:|||||:|
DB 301 ADAASLETNNRITAYTAVTMEISVAVLCPGRLOLDKAVENPEAGQYMGINVTFTSSQT 360
OY 361 L 361
DB 361 L 361

RESULT 2
OY 047115 PRELIMINARY; PRT; 363 AA.
AC 047115;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, last annotation update)
DE COOD PRECURSOR.
GN COOD.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94344028; PubMed=7915003;
RA Froehlich B., Karakashian A., Melsen L.R., Wakefield J.C., Scott J.R.;
RT "Cooc and Cooc are required for assembly of CSI pill."
RL Mol. Microbiol. 12:387-401(1994).
DR EMBL: X76908; CAA54230.1;
DR InterPro: IPR001412; TRNA-synt.L.
DR PROSITE: PS00178; AA_TRNA_LIGASE_I; UNKNOWN_1.
KW SIGNAL.
FT SIGNAL 1 18 POTENTIAL.
FT CHAIN 19 363 COOD.
SQ SEQUENCE 363 AA; 40139 MW; AA9488AADD5F72 CRC64;

```

Query Match 50.3%; Score 949.5; DB 2; Length 363;
 Best Local Similarity 53.1%; Pred. No. 6.4e-66;
 Matches 197; Conservative 55; Mismatches 98; Indels 21; Gaps 11;

```

OY 3 KIIFITFLFSSVLTFAVSADKIPGDE--STINIF-GPR-DRRESSPKNINLNNHITAY 58
    |||:|||||:|
DB 2 KIIFITFLFSSVLTFAVSADKIPGDE--STINIF-GPR-DRRESSPKNINLNNHITAY 57
OY 59 SESHTLYDRMTFLCLSSHNTLNGACPTSENPSSSVS--GETNITLOTEKRSILKRELQI 117
    |||:|||||:|
DB 58 SLHSILTDRIVFLCTSSSNVNGACPTI---GTSQVOYGTITITLOTEKRSILKRLINL 114
OY 118 KGYKOLLFKSVNCG--LTLSAHEFNCNKNA--ASGASLYLYIPAGELKMLPFGGIWDAT 174
    |||:|||||:|
DB 115 AGKKKPLWENOSCDFSNMLVNSKSMSCGAHNGANGLTLLMLYIPAGELKMLPFGGIWEAT 174
OY 175 LKLRVRRYSET---YGTITNITIKLTDKGNIOIWLPOFKSDARVDLNRPTGGTYI 230
    |||:|||||:|
DB 175 LKLRVRRYSET---YGTITNITIKLTDKGNIOIWLPOFKSDARVDLNRPTGGTYI 230
OY 231 GRNSVDMCFYDGYSTNSSLEIRFODNNPKSDGKFYLRKINDDTKEIAYTLLSLAGKSL 290
    |||:|||||:|
DB 234 GSNSLDMCFYDGYSTNSSLEIRFODNNPKSDGKFYLRKINDDTKEIAYTLLSLAGKSL 292
OY 291 TTPNGSTSLNADASLETNNRITAYTMEISVAVLCPGRLOLDKAVENPEAGQYMGINT 350
    |||:|||||:|
DB 293 YPVGOSFTINDSSVLETNNRITAYTAVTMEISVAVLCPGRLOLDKAVENPEAGQYMGINT 352

```

```

OY 351 NYTFPPSSQTL 361
    |||:|||||:|
DB 353 YTFPPSSQTL 363

```

```

RESULT 3
OY 047119 PRELIMINARY; PRT; 364 AA.
AC 047119;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, last sequence update)
DT 01-NOV-1998 (Tremblrel. 08, last annotation update)
DE COOD PRECURSOR.
GN COOD.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=C91F-6;
RX MEDLINE=96071908; PubMed=7591145;
RA Froehlich B.J., Karakashian A., Sakellaris H., Scott J.R.;
RT "Genes for CS2 pill of enterotoxigenic Escherichia coli and their
interchangeability with those for CSI pill."
RL Infect. Immun. 63:4849-4856(1995).
DR EMBL: Z47800; CAA87763.1;
KW SIGNAL.
FT SIGNAL 1 18 POTENTIAL.
FT CHAIN 19 364 COOD.
SQ SEQUENCE 364 AA; 40979 MW; EC7AC738A0E195B6 CRC64;

```

Query Match 49.3%; Score 930.5; DB 2; Length 364;
 Best Local Similarity 48.4%; Pred. No. 1.9e-64;
 Matches 177; Conservative 64; Mismatches 118; Indels 7; Gaps 3;

```

OY 1 MNKILFTFLFSSVLTFAVSADKIPGDE--STINIF-GPR-DRRESSPKNINLNNHITAY 58
    |||:|||||:|
DB 1 MNKILFTFLFSSVLTFAVSADKIPGDE--STINIF-GPR-DRRESSPKNINLNNHITAY 57
OY 59 SESHTLYDRMTFLCLSSHNTLNGACPTSENPSSSVS--GETNITLOTEKRSILKRELQI 118
    |||:|||||:|
DB 61 NKDHSLEDRMTFLCLSSHNTLNGACPTSENPSSSVS--GETNITLOTEKRSILKRELQI 118
OY 119 GYKOLLFKSVNCG--LTLSAHEFNCNKNA--ASGASLYLYIPAGELKMLPFGGIWDAT 175
    |||:|||||:|
DB 119 GYKOLLFKSVNCG--LTLSAHEFNCNKNA--ASGASLYLYIPAGELKMLPFGGIWDAT 178
OY 176 KLRVRRYSETGYTINTITIKLTDKGNIOIWLPOFKSDARVDLNRPTGGTYIGRNSV 235
    |||:|||||:|
DB 179 ELKRRYRYSETGYTINTITIKLTDKGNIOIWLPOFKSDARVDLNRPTGGTYIGRNSV 238
OY 236 DMCFYDGYSTNSSLEIRFODNNPKSDGKFYLRKINDDTKEIAYTLLSLAGKSLTPNG 295
    |||:|||||:|
DB 239 EMLCYDGYSTNSSLEIRFODNNPKSDGKFYLRKINDDTKEIAYTLLSLAGKSLTPNG 298
OY 296 TSLNADASLETNNRITAYTMEISVAVLCPGRLOLDKAVENPEAGQYMGINVTFT 355
    |||:|||||:|
DB 299 EAFITNDSSLEIFNNRIRKSVLPQISIPVLCWPAWLTSEINPEAGEYSGILNVTF 358
OY 356 PSSQTL 361
DB 359 PSSSSL 364

RESULT 4
OY 09XDS1 PRELIMINARY; PRT; 359 AA.
AC 09XDS1;
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-NOV-1999 (Tremblrel. 12, last sequence update)

```


Query Match	Best Local Similarity	5.8%	Score 110;	DB 2;	Length 846;
Matches	86;	Conservative	51;	Mismatches 132;	Indels 124;
				Gaps	20;
QY	42	NESSPKHILNNH1A-----YSESHLYDRMFLCLS--SHNTLNGAC--PTSENP	90		
DB	27	NETLSSDSYNGNVTSDSEFEVKEETSGALITCEGNVCISYACKDPLNKSCEFTE	86		
QY	91	SSSVGGER---NTLQTEKRSI---IKRELQIKGYQLLFKSYNCPGSLTLNSA---	139		
DB	87	FLG-NGYTLCEPDNITQSSHPCALISVSGTNKRLDISGF-SLPSCAYCCPGTGYGAIOT	144		
QY	140	-----HFNCH-----KMAAGASLYIYIPAGELKNLPF-----	167		
DB	145	KGTTTLKDNSSLVLFKNCSTAEAGAIQCKSSSTIELKI---ENKNKLVSENSSK	200		
QY	168	GGIMDATLKL-----RVKRRYSERYTYTINITIKLTDKGNIDIMVPOKSPDAR	216		
DB	201	GAIYADKLTIVYGGPTLFSNNNSVSHNSSPKGA---ICIKNSDG---ECSLTANLGDIT	253		
QY	217	VDLN--LAPTGCGGYIGRSVDM-----CYDGYSTNSSLELTFODN	257		
DB	254	FDGNIIITTTNGSPVTRNSIDLSGGKFKTKLNAKEGFGIFPYDEPIANTGGSTEIEL--N	311		
QY	258	NPKSD---GK--FYLRKINDTKEIAVTL-----SLLACKSLTP	292		
DB	312	KTESPTTYTGKIVFSGEKLSDDEKIVPANLVSYFKOPLKIGAGSLVKDGVLEAKKITO	371		
QY	293	TNGTSLNIADAASLETNNNRITAVTMPETISVPV	325		
DB	372	TKGSTVVMDCITLQTFPSSSGETITLMLDINI	404		
RESULT	7				
ID	P94750	PRELIMINARY;	PRT:	2349 AA.	
AC	P94750:				
DT	01-MAY-1997 (Tremblrel. 03. Created)				
DT	01-MAY-1997 (Tremblrel. 03. Last sequence update)				
DT	01-DEC-2001 (Tremblrel. 19. Last annotation update)				
DE	INVASIN (FRAGMENT).				
OS	Escherichia coli.				
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;				
OX	NCBI_TaxID=562;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-K12;				
RX	MEDLINE=97251358; PubMed=9097040;				
RA	Itch T., Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,				

Query Match	5.8%	Score 109	DB 2	Length 2349
Best Local Similarity	20.0%	Pred. No. 13		
Matches 78	Conservative 65	Mismatches 157	Indels 90	Gaps 15
QY 13	SSVLTTPAVASADKIFGDESTINIFGPRDRNESSPKHNIINNHITAY---SESHLTLYDRMT 69			
DB 746	SAKIATILTSASNGVLANENMAANTVSYNVADGSS---NPINDHTVTFAYVLGSATSEFNMON 802			
QY 70	FLCLSSHHNTLNACPTSENPSSSVSGEINILTLQTEKRSLL-----K 112			
DB 803	---TAKIDVNGLA-TFDLKSSKQEDNVEYVLENGVKOTLLVSVYGDSSSTQVLDLQSK 857			
QY 113	RELQIKGYKQOL-LFKSVNCPFSGLTINTSAHFNCNKNAAASGLYIXYPAGELNLPFGGITW 171			
DB 858	NEVYADGDSVYMTFTVGRKAGNLNDVYVTFNVNVAEAKLSQTEVNSHD-----GIA 910			
QY 172	DATLKLVRKRRYSEYGYTTINITIKLDDKGIQIWLPOFSKDARVDLRLPTGGGYIIG 231			
DB 911	TATLTL-----SLKNGYRVYASVSSGSOANOQVNFIGDOSTAALTLSV-PSGDITV-- 960			
QY 232	RNSVDMCFDYDGSSTNSS---SLEIRFOODNNPKSDGKFYLKRIINDTKRIAYTSLILLAG 287			
DB 961	-----TWTAPOYMTATLQDKNGNLKDKKEITFSPND---VASKFSISNCG 1003			
QY 288	KSLEPTNGTSL-----NIADAASLEIWMNRITAVTMEPISVPVLC 327			
DB 1004	KGMTDSNGVLAISLTLGTLACTHIMARLANSVSDAQPTFPVADNDRAVYVLTQSKAEII 1063			
QY 328	WPG--RIQLDAKVENPEAGQYMGINIVTFT 355			
DB 1064	GNGVDETLTATVKDP-SNHPPAGITVNET 1092			
RESULT 8				
EF76347				
ID P76347	PRELIMINARY	PRT	2383 AA	
AC P76347				
DT 01-FEB-1997 (Tremblrel. 02. Created)				
DT 01-FEB-1997 (Tremblrel. 02. Last sequence update)				
DT 01-DEC-2001 (Tremblrel. 19. Last annotation update)				
DE 02383.				
GN B1978.				
OS Escherichia coli.				
OC Bacteria: Proteobacteria: gamma subdivision: Enterobacteriaceae:				
OC Escherichia.				
OX NCBI_TaxID=562;				
RN [1]				
RP SEQUENCE FROM N.A.				
RC STRAIN=K12;				
RA Blatter F.R., Plunkett G; III, Mayhew G.F., Perna N.T., Glasner F.D.;				
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.				
RN [2]				

RP SEQUENCE OF 2118-2383 FROM N.A.
 RC STRAIN-K12:
 RA Alba H., Baba T., Fujita K., Hayashi K., Honjo A., Horuchi T.,
 RA Ikemoto K., Inada T., Isono K., Isono S., Itoh T., Kanai K., Kasai H.,
 RA Kasimoto K., Kim S., Kimura S., Kitagawa M., Kitakawa M., Makino K.,
 RA Masuda S., Miki T., Mizobuchi K., Mori H., Motomura K., Nakamura Y.,
 RA Nishimoto H., Nishio Y., Oshima T., Saito N., Sempel G., Seki Y.,
 RA Tagami H., Takemoto K., Wada C., Yamamoto Y., Yano M.,
 RL Submitted (FEB-1997) to the EMBL/Genbank/DBD databases.
 DR EMBL: AE000289; AAC/5042.1; -
 DR EMBL: D90837; BAA15800.1; -
 DR InterPro: IPR003344; Big_1.
 DR InterPro: IPR003335; Intimin.
 DR InterPro: IPR002482; LysM.
 DR InterPro: IPR000601; PKD_domain.
 DR Pfam: PF02369; Big_1; 13.
 DR PRINTS: PR01369; INTIMIN.
 DR SMART: SM00257; LysM; 1.
 DR SMART: SM00089; PKD; 4.
 KM Complete proteome.
 SQ SEQUENCE 2383 AA; 251391 MW; 129975F4CC64225C CRC64;

Query Match 5.8%; Score 109; DB 16; Length 2383;
 Best Local Similarity 20.0%; Pred. No. 14;
 Matches 78; Conservative 65; Mismatches 157; Indels 90; Gaps 15;

QY 13 SSVLTFAVSADKIPGDESITNIFGRDRNESSPKHNLNHTAY---SESHLYDRMT 69
 DB 762 SAKIATILSANNNGVLNENANANTSVNVADEGS---NPINDHVTFAVLGSGATSPNNQ 818
 QY 70 FLCLSHNTLNGACPTSENPSSSVSGETNITLQTFEKSLSL-----K 112
 DB 819 ---TAKTGVNGLA--TFDLKSSKQEDNTVEYVLENGVKQTLVSEFGDSSTAQVDLQSK 873
 QY 113 RELQIRGYKQL-LFKSVCNPGSLTNSAHFNCKNANASGSLVLYIPAGELKLNLPFGIM 171
 DB 874 NEVVDAGNDGVMTATVRBAKGNLMDVAVTFVNVNSAEKLSLEVNSHD-----GLA 926
 QY 172 DATLKLVRKRRSEYGYTITNTIKLTDKGNIQIWLPOFKSDARVDLMLRPTGGCTYTG 221
 DB 927 TATLT-----SLKNGDYRVTFASVSSGSOANOQVNFIGQSTALTLSTV--PGDITV-- 976
 QY 232 RNSVDMCFYDGYSTNS-----SLEIFQDNNPKSDGKFLRKINDTKELATYLSLLAG 287
 DB 977 -----TNTAPQYMATLQDKNGNPLKDKETFEFVSPND---VASKFSISNGG 1019
 QY 288 KSLPTNGTSL-----NTADAASLETMNRITAYTAMPETISVPLC 327
 DB 1020 KGMTDNGVAIASLTGTAGTHIMARLANSNVSDAQPMFTFVADKRAVAVLQTSKAEII 1079
 QY 328 WPG--RLQDADAKVNEPDAQYMGMINVTF 355
 DB 1080 GNGVDETTLATATYKDP-SNHPPVAGITVNF 1108

RESULT 9
 Q9RLA0 PRELIMINARY; PRT; 872 AA.
 AC Q9RLA0:
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)
 DE DNA POLYMERASE I.
 GN POLA.
 OS Rickettsia typhi.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Rickettsiae; Rickettsia.
 OX NCBI_Taxid=785;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-WILMINGTON;
 RX MEDLINE=99416441; PubMed=10486973;

RA Andersson J.O., Andersson S.G.E.;
 RT "Genome degradation is an ongoing process in Rickettsia."
 RL Mol. Biol. Evol. 16:1178-1191(1999).
 DR EMBL: AJ238757; CAB56108.1; -
 DR HSSP: P00582; 1KPD.
 DR InterPro: IPR002421; 5.3_exonuclease.
 DR InterPro: IPR002298; DNA_pol.
 DR InterPro: IPR001098; DNA_pol_A.
 DR InterPro: IPR000513; Exo_N1.
 DR InterPro: IPR003584; HHH_2.
 DR Pfam: PF01367; 5.3_exonuclease; 1.
 DR Pfam: PF02739; 5.3_exonuc; 1.
 DR Pfam: PF00476; DNA_pol_A; 1.
 DR PRINTS: PR00868; DNAPOL1.
 DR SMART: SM00475; 5EXOC; 1.
 DR SMART: SM00279; HNH2; 1.
 DR SMART: SM00482; POLA; 1.
 DR PROSITE: PS00447; DNA_POLYMERASE_A; 1.
 SQ SEQUENCE 872 AA; 99309 MW; BC2EBE3490879A08 CRC64;

Query Match 5.7%; Score 108; DB 2; Length 872;
 Best Local Similarity 20.4%; Pred. No. 4.6;
 Matches 77; Conservative 50; Mismatches 139; Indels 112; Gaps 15;

QY 2 NKILFFTLFFSSVLTFAVSADKIPGDESI-----TNIIFGRDRNESSPKH 48
 DB 383 NKSIRKITYSLKHLLEFVANSQSHKRTAIEDLMOYTLNAGLVQKNLFA-----ETLTKD 437
 QY 49 NILNNHTAYSESHLYDRMTFLCLSSHNTLN-----GACPTSENPSSSVSGETNIT 101
 DB 438 NIINSEAKYINFTISLY-KOTLLALQKKAFLRYREIDLPCTCIIDMKETIGIKVANYI 496
 QY 102 LQTFEKSLSL-----KCYKQLFKSVNCPGSLTNSAHFNCKN 146
 DB 497 HQLSDFGEFELIKIEEIFALSGTKFNINASQQLSEILFKMKQLPQGNFLAKTSYSTK- 555
 QY 147 AASGASLYLYIPAGELKLNLPFGIMATLKL-----VKRRSEYGYTITNTIKL-- 198
 DB 556 -----AGILKLSSEDEYHATLTLRMRQLTKNTYDLSLPROQINNTKRIHT 603
 QY 199 -----TDKGNIQIWLPOFKSDARVDLMLRPTGGCTYIGNSVDMCFY--DGY---STNS 247
 DB 604 TFLQSTTTTGRSLSSQEPNLD-----NIPTRSSG-----NKRLQAFIADGKILSADY 652
 QY 248 SLEIRF-----QDNNPKSDGKFLRKINDTKE-----IAY 279
 DB 653 SQELRLILSHIANVDYLKQAFINKEDIHQTAQIFNLQKHELTSERRKAKAINFGIIV 712
 QY 280 TISLLAGKSLPTNGTS 297
 DB 713 GISAFGLAQNLVNSGTA 730

RESULT 10
 O25579 PRELIMINARY; PRT; 2529 AA.
 AC O25579:
 DT 01-JAN-1998 (TREMBLrel. 05, Created)
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE TOXIN-LIKE OUTER MEMBRANE PROTEIN.
 GN HP0922.
 OS Helicobacter pylori (Campylobacter pylori).
 OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
 OC Helicobacter.
 OX NCBI_Taxid=210;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=26695 / ATCC 700392;
 RX MEDLINE=97394467; PubMed=9252185;
 RA Tomb J.-F., White O., Kertlavage A.R., Clayton R.A., Sutton G.G.,
 Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,

RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Lotius B., Richardson D., Dodson R., Khalak H.G., Glodok A.,
RA McEnany K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Goodyear J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weiman J.M., Fujii C., Bowman C., Matthey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Kaip P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen *Helicobacter*
RT *pylori*.";
RL Nature 388:539-547(1997).
DR EMBL: AE000602; AAC07969.1; -
DR TIGR: HP0922; -
DR InterPro: IPR004311; VACA.
DR Pfam: PF03077; VACA.3.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 2529 AA; 274561 MW; 440882E8644472EC CRC64;

Query Match 5.7%; Score 108; DB 16; Length 2529;
Best Local Similarity 22.5%; Pred. No. 17;
Matches 84; Conservative 50; Mismatches 122; Indels 118; Gaps 22;

OY 85 TSENSSSSVSEETITLOFTKRSLL-----KRELOIKGYKOLLER---SVNCP 131
DB 977 TSNFATTTQLLGNTMTTL---SSQSLNFNGDTTLQNNANITLGNKSCAFAFNLSLTLDNN 1033
OY 132 SGLTL-NSAHFNCNKNA--SGASLYLYPA-GELKNLPFGGIMDATLKVRRRYSEY 187
DB 1034 SNLSLNDQSVLANNTSAFNNOASLNIYNGSOATNSLFPNG---GTLSLNASSKINASN 1090
OY 188 GYTYITITIKLTD-----KGNIOIWLPOKSDARVDLNLPTGGGTYIGRNSVD 236
DB 1091 ASFSNNTTINLDDSVLASNTSSLVNANI---NFGASQADF---CGNTITIDTASFN 1140
OY 237 MCFYGGYGTINSSSL---EIRFDNNPK-----SDGKFLRKIN---DDT 274
DB 1141 --FDSASSLNFNMLANGALNNGYTPSLTKALMSVSGFVLGNNGDMLSDINIFDNT 1198
OY 275 KEIAYTLLSLLAGKSLTPTNGTS-----LNTADAASLEFN---WNRITAVTMPET 321
DB 1199 KSVY--NLTNOKGKITGSGANGYEKILFYGMKTIOMATYSNNNNIQTWFSFNLNSSOI 1256
OY 322 -----SVPVL-----CWPGRLOLDAYENP-----EAGGY--M 347
DB 1257 IQESIKNGDLTIEVLNPNNSASNTIFNIPELNYQASKQNPFGSYDSDMAGTYVLT 1316
OY 348 GNINVTFTF-SSQT 360
DB 1317 SNIKGLFTPKGSQT 1330

RESULT 11

O96VN2 PRELIMINARY: PRT; 430 AA.
AC O96VN2;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE EXTRACELLULAR ALKALINE PROTEASE.
GN Blumeria graminis.
OS Blumeria graminis.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Leotiomycetes;
OC Erysiphales; Erysiphaceae; Blumeria.
OX NCBI_TaxID=34373;
RN [1]
RP SEQUENCE FROM N.A.
RA Zhang Z., Gurr S.J.;
RT "An extracellular alkaline protease from *Blumeria graminis*.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY039006; AAK84436.1; -
KW protease.
SQ SEQUENCE 430 AA; 44817 MW; 8984D822BBBC3CB5 CRC64;

Query Match 5.6%; Score 105.5; DB 3; Length 430;
Best Local Similarity 21.7%; Pred. No. 3;
Matches 86; Conservative 48; Mismatches 128; Indels 135; Gaps 21;

OY 7 ITTFFSSVLTFFAVSADKI---PGDESITNIFSPDRNESPKNH--ILNNHTIAYS-- 59
DB 12 VITLY---LFSFAYSAPPLFSPANPPSIYGLDLPISNLDPAEDMIANRYIVVYSKD 67
OY 60 -----ESHITLDRMTPLCLSSHNTLN-----GACPTSENSS 91
DB 68 ATDDAVESH---QSTIKTALKKSLNATSIDGRLSPMHTFKKGRGMOCLDEDMI 123
OY 92 SSVSEETITLOFTKRSLLIKRELOIKGYKOLLERFVSNCPGSLT-LSAHFNCNK----- 145
DB 124 IEIEASEV-----SVYEDATQV-GPLELT-EQIEAPIGLVRLSHALSESSEYFED 173
OY 146 NASGASLYLY-IPAGELKNLPFGGIMDATLKVRRRYSEYTYTINITIKLTDKNT 204
DB 174 NASDAGIYGYIDPGRASHFEG-----RATMKAN- 205
OY 205 QTWLPOKSDARVDLNLPTGGGTYIGRNSVDMCFYDGTNSSSLEIRFDNNPKSDG 264
DB 206 -----FTDDINEDON---GHSVAVTTIGASF--GVAKITELVGIKVLDAKKGSA 253
OY 265 FYLRKITNDTKEIAYTLLSLLAGKSLTPTNGSLNIAADAASLETNMNRITAVTMPETISVP 324
DB 254 NVLRGVNPDVNERG-----LAGKAV-----VNISIGSKSKALN-TAIEALTAKGYTVV 301
OY 325 VICWPGRLQLDKAVENPEAGQYMNIN--VTFPSS 358
DB 302 VAA-----GNSNKDATSFSPAS 318

RESULT 12

O96FL1 PRELIMINARY: PRT; 1441 AA.
AC O96FL1;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)
DE UNKNOWN PROTEIN.
GN Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Lactococcus.
OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IL1403;
RX MEDLINE=21235186; PubMed=11337471;
RA Bolotin A., Wincker P., Mueger S., Jallion O., Melaire K.,
RA Weissenbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium *Lactococcus*
RT *lactis* ssp. *lactis* IL1403.";
RL Genome Res. 11:731-753(2001).
DR EMBL: AE006376; AAK05556.1; -
DR InterPro: IPR001899; Gram_pos_anchor.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 1441 AA; 142946 MW; 4CD4ABC561BA15EA CRC64;

Query Match 5.5%; Score 104.5; DB 16; Length 1441;
Best Local Similarity 21.4%; Pred. No. 16;
Matches 91; Conservative 59; Mismatches 160; Indels 115; Gaps 21;

OY 10 IFFSVSLFTF-----AVSADKI-PG-----DESTINIFGPRDRNESPKNH----- 50
DB 28 LYASSVLLIYVGAIAAPSVKADYLPQGQGTGANTATGALGSSAGSAGSNNVSLAN 87
OY 51 -LNHTITAVSESHLYLDRMTPLCLSSHNT--LNGACPTSENPS----- 91

Db 88 SANN--VNYSSAOSAASSVYSSSSAASSNPLSANDPSSYTDGSSVNSNSQLKT 145
QY 92 -----SSVSGEINT-----LQTEKRLIKRELQIKGKOLLFKSVNCPGLT 135
Db 146 DSGVENSATNTNSYVODLNSSGAGTATGSSAQSASLATPGSSDALKT-----SDTN 201
QY 136 LNSAHFNCKNAAGSASLYLYIPAGELKNLPGGIMDATLKLVRKRYSEFYGTINT 195
Db 202 LNS---NGAVTAGSGASSVSFDIKDIOVPGA-----LNIYNGATGSI-TWTVSNP 249
QY 196 IKLTDKG-----NIQIMLPQFKSDARVDLN---LRPGGCTY-IGRNSVDMCFYDGYSTN 246
Db 250 SDLTNAGLSVANVAVYDASKGDGALNVNGNEAYYQTAGGDVYIGTENY----PGYPTG 304
QY 247 SSSLEIRQDNNPK-----SDGKFYLRKINDPQKEIAYTSLSLAGSLTPT 293
Db 305 SSASMEPTTESASIMASASSVAASYSIDANIALASOYADAGSL-YNSYAAANGDGTAA 363
QY 294 NCTSLINADASLETNNRITAVTPEISVPLVCPGRLOLDKAKENPEAGGYMGINVT 353
Db 364 SYANLVVAIQASISANNALALDGT-----PGYLR-----SSVASQAV--VNT 408
QY 354 FTPSS 358
Db 409 YTTDS 413

RESULT 13
QY 09A005 PRELIMINARY: PRT: 803 AA.
AC 09A005:
DT 01-JUN-2001 (TREMBLREL. 17, Created)
DT 01-JUN-2001 (TREMBLREL. 17, Last sequence update)
DT 01-DEC-2001 (TREMBLREL. 19, Last annotation update)
DE HYPOTHEITICAL 89.2 KDA PROTEIN.
GN OSJNBA0040E01.17.
OS Oryza sativa (Rice).
OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta;
OC Spermatophyta: Magnoliophyta: Liliopsida: Poales; Poaceae;
OC Erihaldoideae: Oryzae; Oryza.
OX NCBI_Taxid=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPONBARE;
RA Bueli C.R., Yuan Q., Ouyang S., Moffat K.S., Hill J.N., Burr P.C.,
RA Hsiao J., Zismann V., Pal G., Bowman C.L., Fujii C.Y., Vanaaken S.E.,
RA Bowman C.L., Craven B., Uteback T.R., Khalak H., Feldlyum T.V.,
RA Oryza sativa chromosome 3 BAC OSJNBA0040E01 genomic sequence.*;
RT Submitted (AUG-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL: AC079887; AAK16187.1; -.
KM Hypothetical protein.
SQ SEQUENCE 803 AA; 89209 MW; E119841096A8BDAC CRC64;

Query Match 5.5%; Score 104; DB 10; Length 803;
Best Local Similarity 20.2%; Pred. NO. 8.5; Indels 68; Gaps 10;
Matches 65; Conservative 44; Mismatches 145;

QY 33 TNIF-GPRDRNESSPKH-----NILNNH-ITWYSESHLYDRMTFLCLSSH 76
Db 406 TNIIDGOSLTKTYISHDGLVATNGVKGRDVMENHPLGYSSSHAHHPQWELPMCTGS 465
QY 77 NTLGACCTSENSSSVSGETNITLQTEKRLIKRELQIKGYKOLLFKSVNCPGSLTL 136
Db 466 HLLPPLPMISNNPNSSFRGTGLKAEILMGAYSTENTENLIDNRR--KGIIKPKSSQ 522
QY 137 NSAHFNCKNAAGSASLYLYIPAGEL-----KNLPRGGT-----WDATIKRVRKRYSE 185
Db 523 NVAENICKAAEYOCNDISQIVCNOOYVLLPLNKASHFGGLPTKKFEDKLVSROKKRRA 582
QY 186 TYGVTITITIKLDKGNIQIMLPQFKSD-ARVDLNLRTPTGGCTYIGNSVDMCFYDGY 244

Db 583 T-----SLLSWHAQVMGSGSKTHHTRKRELDMAHARTRLVEKVDANMK 626
QY 245 TSSSLEIRQDNNPKSDGKFYLRKINDPQKEIATLSSLAGKSLTPNGSLINADAA 304
Db 627 TKNSTF-----VSOAQKRLAFTTKIMQYILPDLPRLLAANAID-- 665
QY 305 SLETNNRITAVTPEISVPL 326
Db 666 SCETIYKRSRLALPDARNPAT 687

RESULT 14
QY 09RZS7 PRELIMINARY: PRT: 691 AA.
AC 09RZS7:
DT 01-MAY-2000 (TREMBLREL. 13, Created)
DT 01-MAY-2000 (TREMBLREL. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLREL. 17, Last annotation update)
DE HYPOTHEITICAL 69.7 KDA PROTEIN.
GN DRB0037.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
OX NCBI_Taxid=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RI.
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Uteback T., Zaleski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
radiodurans R1."
RL Science 286:1571-1577(1999).
DR EMBL: AE001826; AAF12628.1; -.
DR TIGR: DRB0037; -.
DR InterPro: IPR002048; EF-hand.
DR InterPro: IPR001899; Gram_pos_anchor.
DR PROSITE: PS00018; EF-hand. UNKNOWN.1.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN.1.
KM Hypothetical protein; Plasmid; Complete proteome.
SQ SEQUENCE 691 AA; 69702 MW; 744112A6A23C022E CRC64;

Query Match 5.5%; Score 103.5; DB 16; Length 691;
Best Local Similarity 21.3%; Pred. NO. 7.7; Indels 97; Gaps 15;
Matches 70; Conservative 40; Mismatches 121;

QY 28 GDESTINIFGPRDRNESSPKHILNNHITAVSESHLYDRMTFLCLSSHNTLNGACPTSE 87
Db 218 GSPILTNTVVPSSRPRLPPTG-----TTPGTGTGVSNGN 253
QY 88 NESSSVSGEINTI-TLQTEKRLIKRELQIKGYKOLLFKSVNCPGSLTNSAHFNCKN 146
Db 254 GPAESTGPAVYTGIGSTGTOPPVVDASGNOYAYPKAAEFVN-PDSVMTIITINPPT 312
QY 147 AASGASLYLYIPAGELKNLPGGIMDATLKLVRKRYSEFYGTINTIKITD-KGNIQ 205
Db 313 AAT-----YELV--DLGSLPAG-----TVVFTDANGNP- 340
QY 206 IWLPOFKSDARVDLNLRTPGG-GTYIGRNSVDMCFYDGYSTNSSLERFQ--DNPKS 261
Db 341 -LPDTDNGREYVAGNGGTATY-----RVVYVTPDRESAAAVAPRIRIPGVGDN--- 391
QY 262 DGKFLRKINDPQKEIATLSSLAGKSLTPNGSLINADASLETNNRITAVTPEI 321
Db 392 -----RGIDVATYTNVLLSNLKEFGNTGNTALGVSDVPVRYT-----VTOPSV 435
QY 322 SVPLVCPWGRLOLDKAKENPEAGGYMGN 349


```

51 TACATTGGCTGTATCGCAGATAAATTCGCCGAGATGAAGCATATCA 100
   :      :      :      :      :      :      :      :
17 rglYlnserThrphIsthrasnValGlnAlaGlySerIleasnLysThr 34
   :      :      :      :      :      :      :      :
101 AATATTTTGGCCCGCGTGACGAGAAC.....GAATCTTCCCCCAACAT 144
   :      :      :      :      :      :      :      :
34 lnsertIleGlyProIleasprGserAlaAlaIleSerTyProAlaHis 50
   :      :      :      :      :      :      :      :
145 AATATTTTAAATACCATATTTACAGATACAGGAAAGTCATCTGTGA 194
   :      :      :      :      :      :      :      :
51 TyIlePheHisGlnHisValAlaGlyTyAsnLysAspHisSerLeuPh 67
   :      :      :      :      :      :      :      :
195 TGAATAGATGACTTTTATGTTGTCTTCTCACAATACACTTAATGAG 244
   :      :      :      :      :      :      :      :
67 eAsparGmetThrPheLeuLysMetSerSerThrAspAlaSerLysGly 84
   :      :      :      :      :      :      :      :
245 CATGTCACCAACGATGAGAAATCTCAGCATTCATCGCTCAGCGGTGA 294
   :      :      :      :      :      :      :      :
84 lAcysProThrGlyGlnAsnSerLysSerSer.....GlnGlyGlnThr 98
   :      :      :      :      :      :      :      :
295 AATATTAACATTAACATTTACGGAAGAAAGAGTTAATAAAGACAGCT 344
   :      :      :      :      :      :      :      :
99 AsnIleLysLeuIlePheThrGlnLysLysSerLeuAlaArgLysThrI 115
   :      :      :      :      :      :      :      :
345 ACAAAATTAAGGCTATTAACAAATTTATGTTCAAAAGTGAATGTC... 390
   :      :      :      :      :      :      :      :
115 uAsnLeuLysGlyTyLysArgPheLeuTyrgLuserAsparGcysIle 132
   :      :      :      :      :      :      :      :
391 .....CCATCCGGCCTAACACTTAACTCAGCTCATTTTAACTGTAATA 435
   :      :      :      :      :      :      :      :
132 lStyValAspLysMetAsnLeuAsnSerHisThrValLysCysValGly 148
   :      :      :      :      :      :      :      :
436 AAGCGGCTCAGGTGAAGTTATATTTATATTCCTCCGCGGACAGCT 485
   :      :      :      :      :      :      :      :
149 SerPheThrArgGlyValAspPheThrLeuTyIleProGlnGlyGln 165
   :      :      :      :      :      :      :      :
486 AAAAAATTTGCCCTTTGGTGTATCTGGAGTCTACTCTGAAGTGAAG 535
   :      :      :      :      :      :      :      :
165 eAspGlyLeuLeuThrGlyGlyIleThrGlnAlaThrLeuGlnLeuArg 182
   :      :      :      :      :      :      :      :
536 TAAAAACAGCATATAGTACAGCTATGGAACCTTACACTATTAATATCA 585
   :      :      :      :      :      :      :      :
182 AllysArgHisTyAspTyAsnHisGlyThrTyLysValAsnIleThr 198
   :      :      :      :      :      :      :      :
586 ATTAATTAAGTATGATGAGGAATATTCAGATATGTTACCTGACCTCA 635
   :      :      :      :      :      :      :      :
199 ValAspLeuThrAspLysGlnAlaGlnValIleThrProLysPheH 215
   :      :      :      :      :      :      :      :
636 AAGTGAAGCTCGCGTGCATCTTAACCTGCTCCACCTGCTGGGCGACAT 685
   :      :      :      :      :      :      :      :
215 sSerAspProArgIleAspLeuAsnLeuArgProGlnGlyAsnGlyLys 232
   :      :      :      :      :      :      :      :
686 AATTTGGAAGAAATTTCTGTTGATATGCTTTTATGATGATATAGTACT 735
   :      :      :      :      :      :      :      :
232 ySerGlySerAsnValLeuGlnMetCysLeuTyAspGlyTySerThr 248
   :      :      :      :      :      :      :      :
736 AACAGAGAGCTCTTTGGAGATTAAGATTTCCAGATTAACAATCCCTAA 785
   :      :      :      :      :      :      :      :
249 HisSerGlnSerIleGlnMetArgPheGlnAspAspSerGlnThrLys 265
   :      :      :      :      :      :      :      :
786 TGGGAATTTTATCTAAGGAAATAAATGATGACACCAAGAAATTCGAT 835
   :      :      :      :      :      :      :      :
265 nAsnGlnTyAsnLeuIleLysThrGlyGlnProLeuLysLysLeuPro 282
   :      :      :      :      :      :      :      :
836 AATCTTTGACCTTCTCTGGCGGGTAAAGTTTAACTCCACCAAAATGGA 885
   :      :      :      :      :      :      :      :
282 yLysLeuSerLeuLeuLeuGlyArgGlnPheTyProAsnAsnGly 298
   :      :      :      :      :      :      :      :
886 AGCTATTAATTAATTTGCTGACGAGAGCTTCTCGGAACAACAACGAAAT 935
   :      :      :      :      :      :      :      :
299 GlnAlaPheThrIleAsnAspThrSerSerLeuPheIleAsnTrpAsn 315
   :      :      :      :      :      :      :      :
936 AATTACAGCTGTACCATGCCAGAAATCAGTTCGCGTCTGTGTGGC 985

```

```

||||| .....|.....|.....|.....|.....|.....|.....|.....|
315 gIleLysSerValSerLeuProGlnIleSerIleProValLeuCysTrp 332
   :      :      :      :      :      :      :      :
986 CTGGACGTTTGCATTTGATGCAAAAGTGAAATCCGAGCTGACAA 1035
   :      :      :      :      :      :      :      :
332 roLAsnLeuThrPheMetSerGlnLeuAsnAsnProGlnAlaGlyGlu 348
   :      :      :      :      :      :      :      :
1036 TATATGGTAAATTAATGTTACTTTCACACCAAGTAGTCAAACTGTC 1083
   :      :      :      :      :      :      :      :
349 TySerGlyIleLeuAsnValThrPheThrProSerSerSerLeu 364
   :      :      :      :      :      :      :      :
seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT: AAB45919
seq_documentation_block:
ID AAB45919 standard; Protein; 359 AA.
XX
AC AAB45919;
XX
DT 23-MAR-2001 (first entry)
XX
DE S. enterica serovar Typhi, tcfD fimbrial subunit protein.
XX
KW Fimbrial protein; saf; tcf; vaccine; gene therapy; immunization;
XX tcf insert; detection.
XX
OS Salmomella typhi.
XX
PM WO200073336-A1.
XX
PD 07-DEC-2000.
XX
PF 26-MAY-2000; 2000WO-SE01079.
XX
PR 28-MAY-1999; 99SE-0001961.
XX
PA (ACTI-) ACTIVE BIOTECH AB.
XX
PI Folkesson A, Normark S, Loefdaahl S;
XX WPI: 2001-061512/07.
XX DR N-PSDB: AAC82926.
XX
PT Fimbriae proteins of Salmomella enterica subspecies I bacteria, useful
PT for producing vaccines against the bacterial subspecies and for
PT detecting the bacteria
XX
PS Disclosure: Page 72-73; 77pp; English.
XX
CC This invention describes the novel proteins (saf and tcf) (I) encoded by
CC a DNA sequence of a gene encoding the precursor of the saf fimbriae unit
CC of Salmomella enterica subspecies I or a DNA sequence of the gene
CC encoding the tcf fimbriae unit of S. enterica subspecies I serovar Typhi
CC The products of the invention can be used as vaccines or for gene
CC therapy. Such vaccines are useful for protection against diseases caused
CC by S. enterica subspecies I or S. enterica subspecies I serovar Typhi.
CC The saf and tcf proteins from S. enterica subspecies I bacteria are
CC useful for active or passive immunization in mammals. The nucleotide
CC sequences are useful for constructing vectors for use as vaccines for
CC insertion into attenuated bacteria in constructing a recombinant viral
CC vaccine, or for direct inoculation of a nucleic acid vaccine. The protein
CC or antigenic fragments, nucleic acid sequences, and antibodies are useful
CC in molecular diagnostic assays for the detection of S. enterica
CC subspecies I.
XX
SQ Sequence 359 AA;
XX
alignment_scores:
Quality: 275.50 Length: 303
Ratio: 1.514 Gaps: 11
Percent Similarity: 60.066 Percent Identity: 29.703
alignment_block:

```


US-09-839-894-9 x AAB45919

Align seg 1/1 to: AAB45919 from: 1 to: 359

```
208 TTTTATGCTGTTCTCTGCACATACACTTAATGAGCATGTCACACAG 257
      ::::: ||| |||::: |||::: |||::: |||::: |||:::
79 TrpValCysArgSerIsmrIasnIugIyAlaIacysGIuGIuTh 95
      ::::: |||::: |||::: |||::: |||::: |||::: |||:::
258 TGAAGATCCATGACAGTTCATCGGTCGCGGTAACAATTAACATTAC 307
      ::::: |||::: |||::: |||::: |||::: |||::: |||:::
95 rHisLeuValITrPTryAlaPheGlyAlaIysSerIysIleArgLeuA 112
      ::::: |||::: |||::: |||::: |||::: |||::: |||:::
308 AATTTCGGGAAAAAGAACTTTAATAAAAAGAGACTACAAATTAAAGC 357
      ::::: |||::: |||::: |||::: |||::: |||::: |||:::
112 rGpHeArgIuGlnIleSerHisAlaGlnIlePheLeu..... 124
      ::::: |||::: |||::: |||::: |||::: |||::: |||:::
358 TATTAACAATTATATGTTCCAAAGTGTAAAC.....TGCCCATCGGCGCT 401
      ::::: |||::: |||::: |||::: |||::: |||::: |||:::
125 .....IleuLeuGlySerValAlaArgAspAlaCysTyrThrGlyVa 138
      ::::: |||::: |||::: |||::: |||::: |||::: |||:::
402 AACACATTAACTGACGTCATTTTAACTGTAAATAAAAAGCGGCTTCA... 447
      ::::: |||::: |||::: |||::: |||::: |||::: |||:::
138 L.....IleasnMetasnAlaAlaIacysG 147
      ::::: |||::: |||::: |||::: |||::: |||::: |||:::
448 ....GTCGCACGTTATATTTATATATTCCTGCTGGCGCACTAAAAAT 492
      ::::: |||::: |||::: |||::: |||::: |||::: |||:::
147 IntPrpIAlgSerIleuLysLeuArgIleProSerGIuGIuLeuAlaIys 163
      ::::: |||::: |||::: |||::: |||::: |||::: |||:::
493 TTGCGCTTTGGTGCTATCTGGGATCGGTACTGTGAAGTTA..AGAGTAA 539
      ::::: |||::: |||::: |||::: |||::: |||::: |||:::
164 IleProThSerGIyThrTPrIysAlaThrLeuValIleuAspTyrLeuGI 180
      ::::: |||::: |||::: |||::: |||::: |||::: |||:::
540 AAGACGATATAGAGAACCTTATGAACTTACACTTAAATATACATATTA 589
      ::::: |||::: |||::: |||::: |||::: |||::: |||:::
180 nTrpGIyGIaAspAspProLeuGIyThrSerThrIhrAspIleThrIleuA 197
      ::::: |||::: |||::: |||::: |||::: |||::: |||:::
590 AATTAACTGAT.....AAGGCAATATTCAGATATGTTACCTCAGTTC 633
      ::::: |||::: |||::: |||::: |||::: |||::: |||:::
197 snValIhrAspHisPheAlaGlnAsnAlaAlaIleTyrPheProGIuPhe 213
      ::::: |||::: |||::: |||::: |||::: |||::: |||:::
634 ...AAAAGAGCGGCTGCGCTGCATGCTTAACTTGCGGCCAACGTGGGGG 680
      ::::: |||::: |||::: |||::: |||::: |||::: |||:::
214 GIYThAlaIThrProIArgValAspLeuAsnIleuHisArgMetAsnAlaSe 230
      ::::: |||::: |||::: |||::: |||::: |||::: |||:::
681 CACATATATTTGGAAGAAATCTGTGATATGCTTTATGATGATGATATA 730
      ::::: |||::: |||::: |||::: |||::: |||::: |||:::
230 rGlnMetSerGIyArgAlaAsnLeuAspMetCysLeuIyrAspGIy...G 246
      ::::: |||::: |||::: |||::: |||::: |||::: |||:::
731 GTACTAAACAGCAGCTCTTTGAGATPAAGATTTTCAGATPAACAACTGAA 780
      ::::: |||::: |||::: |||::: |||::: |||::: |||:::
246 IyAlaIysAlaArgSerIleuGIuMetIysIleGIuGIySerAsnIysSer 262
      ::::: |||::: |||::: |||::: |||::: |||::: |||:::
781 TCGATGAGGGAATTTATATTAAGGAAATTAATGATGACACCAAGAAT 830
      ::::: |||::: |||::: |||::: |||::: |||::: |||:::
263 GIYThGIyPheGlnValIleIysSerAspSerAlaAspTh.....I 277
      ::::: |||::: |||::: |||::: |||::: |||::: |||:::
831 TGCATATACCTTTGTCACCTCTCTGGCGGGTAAAGATTTAACTCCACAA 880
      ::::: |||::: |||::: |||::: |||::: |||::: |||:::
277 eaSpYrAlaValSerMetAsnIyGIyAlaArgSerIleProValIThrA 294
      ::::: |||::: |||::: |||::: |||::: |||::: |||:::
881 ATGGAACGTCATTAATATGCTGACGCGACGCTTCTGTGAAACAACAGTGG 930
      ::::: |||::: |||::: |||::: |||::: |||::: |||:::
294 rGlnIyAlaGIuPheSerIleuAspAsnValaIAspIalAlaIhrArg... 309
      ::::: |||::: |||::: |||::: |||::: |||::: |||:::
931 AATAGAATTAACAGCTGTGCACATGCCAGAAATCAGATGTTCCGGCTGTGG 980
      ::::: |||::: |||::: |||::: |||::: |||::: |||:::
310 .....ProValIleuProGIyGlnArgIlnAlaValaIArgCy 322
      ::::: |||::: |||::: |||::: |||::: |||::: |||:::
981 TTGCGCTGAGAGCTTTGCAATTTGAGATGACAAAA.....GTGGAATATC 1021
      ::::: |||::: |||::: |||::: |||::: |||::: |||:::
322 sValProValProLeuIThrLeuIThrThcIlnProPheAsnIleArgIuL 339
      ::::: |||::: |||::: |||::: |||::: |||::: |||:::
1022 CGAGGCTGACACATATATATGCGTAATATTTAATGTTACTTCACACCAAGT 1071
```

```

.....|::|||::|||||:.....|::|||::|||:..
339 ysaRSeGclYgluYrGlmeIynHlneuvrValThrmleumetGIy 355
          ::|||
          356 ThrGlnThr 358

seq_name: /SDSL/gcgdata/hold-geneseq/geneseg-emb1/AA2000.DAT:AAV50814

seq_documentation_block:
ID   AAV50814 standard; Protein; 770 AA.
XX
AC   AAV50814;
XX
DE   D. melanogaster acetyl-choline receptor protein from clone Da7.
XX
KW   Acetyl-choline receptor; nicotinic insect; insecticide; screening;
XX neurotransmission; plant protection agent; conductance; AChR.
XX
OS   Drosophila melanogaster.
XX
PE   DE|9819829-A1.
XX
PI   11-NOV-1999.
XX
PF   04-MAY-1998; 98DE-1019829.
XX
PR   04-MAR-1998; 98DE-1019829.
XX
PA   (FARB ) BAYER AG.
XX
PT   Adamczewski M, Oellers N, Schulte T;
DR   WPI: 2000-014207/02.
DS   N-PDB; AAZ24475.
XX
XX New nucleic acid encoding a nicotinic acetylcholine receptor from
XX insects, used to identify potential insecticides -
XX
PS Example 1a; Page 12-14; 26pp; German.
XX
CC This invention describes a novel nucleic acid (NA) encoding a nicotinic
CC acetyl-choline receptor (I) from insects which can be used as an
CC insecticide. Inhibitors of (I) interfere with neurotransmission. (I)
CC (also vectors containing it, its regulatory regions, and antibodies
CC directed against (I)-encoded proteins) are used to screen for: (a) plant
CC protection agents that alter conductance of AChR, potentially useful as
CC insecticides, or (b) genes which encode polypeptides that are involved in
CC formation of functionally related AChR in insects. (I) are also used to
CC isolate and characterize the specified regulatory regions and for
CC recombinant production of (II). This sequence represents an
CC acetyl-choline receptor isolated from Drosophila melanogaster.
XX
SQ Sequence 770 AA;

alignment_scores:
Quality: 119.50 Length: 406
Ratio: 0.653 Gaps: 20
Percent Similarity: 45.074 Percent Identity: 20.443

alignment_block:
US-09-839-894-9 x AAV50814 ..

Align seg 1/1 to: AAV50814 from: 1 to: 770

25 ACATTGTTCCTCAGTACTTTTACATTTGGTGCGAGATAA 74
      ||::||| || | ||| ||| :||| |||::|||:
249 ThrlleaIpheIlleserTYrlengllSerPhealalaagInleuylsas 265
'5 AATTCGCCGAGAATAAACATAACTAAATATTTTTGGCCGCCTGACAGGA 124
```

[illegible]

513	etlellelrahrgrhtrhrleuYrYrPhrPheAsnleulleleptoc	530
841TTGTCACCTCTCTTGGCGGATAAAGTTAACTCCACA	879
	: :	
530	yValleullelasermetAlaleuenglYherThrleuProProasp	546
880	AATGGACGTCAATTAATATGCTGCAGCAGCTTCTCTGAA...ACAA	926
	: : : : :	
547	SergIgluLysleuSerleGlyValThrIleleuSerleuThrva	563
927	CTGGAAATGAAATTACAGCTGCATCCAGCCAGAAATCAGT.....GTTC	970
	: : :	
563	IpheleuAsnmetvalAlaaglurhmetProAlaThrSerAspAlaValp	580
971	CGGTG.....TTGTGTGG.....	984
	:	
580	roleuThrIleargIleValIpheleuLysrtrleuProThrIleleuarg	596
985CCTGGACGT 993	
597	MetSerArgProGlyArg 602	
seq_name:	/SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA194.DAT:AA47575	
seq_documentation_block:		
ID	AA47575 standard; Protein; 650 AA.	
AC	AA47575;	
XX		
DT	19-JUL-1994 (first entry)	
XX		
DE	Alpha-agglutinin of <i>Saccharomyces cerevisiae</i> .	
XX		
KW	Immobilisation; enzyme; cell wall; alpha agglutinin; AGA 1; FLO 1;	
KW	Major cell wall protein; glycosyl-phosphatidyl-inositol;	
KW	anchoring protein; alpha factor; alpha-agglutinin; invertase;	
KM	inulinase; alpha-amylase; <i>Saccharomyces cerevisiae</i> ;	
KW	enzymatic process; fermentation; biodegradation; catalysis.	
XX		
OS	<i>Saccharomyces cerevisiae</i> .	
XX		
PN	WO9401567-A.	
PD	20-JAN-1994.	
XX		
PF	07-JUL-1993; 93WO-EP01763.	
XX		
PR	08-JUL-1992; 92EP-0202080.	
PR	14-DEC-1992; 92EP-0203699.	
XX		
PA	(UNIL) UNILEVER NV.	
PA	(UNIL) UNILEVER PLC.	
XX		
PI	Klis FM, Schreuder MP, Tosechka H, Verrips CT;	
XX		
DR	WPI: 1994-035071/04.	
DR	N-PSDB: AAO54012.	
XX		
PT	Immobilisation of enzymes to microbial cell wall - by prodn. of	
PT	fusion protein of enzyme linked to anchoring protein	
XX		
PS	Example 1; Page 32-39; 99p; English.	
XX		

The alpha-agglutinin is used in a method to immobilise enzymes to a microbial cell wall. The coding sequence is used in the production of a recombinant polynucleotide which comprises a structural gene encoding a protein with catalytic activity and at least part of a gene encoding at least the C-terminus of a protein capable of anchoring in a eukaryotic or prokaryotic cell wall. The anchoring fragment or protein is selected from alpha agglutinin, AGA 1, FLO 1, major cell wall protein of lower eukaryotes or a proteinase of lactic acid bacteria. The recombinant polynucleotide preferably also comprises a sequence encoding a signal peptide to ensure

CC secretion of the expressed product. The signal peptide is
 CC preferably derived from glycosyl-phosphatidyl-inositol, anchoring
 CC protein, alpha factor, alpha-agglutinin, invertase or inulinase,
 CC alpha-amylase of Bacillus or proteinases of lactic acid bacteria.
 CC The host microorganism can be used for performing enzymatic
 CC processes on an industrial scale.

XX Sequence 650 AA:

alignment_scores:

Quality: 114.00 Length: 469
 Ratio: 0.573 Gaps: 23
 Percent Similarity: 42.431 Percent Identity: 20.896

alignment_block:

US-09-839-894-9 x AAR47575 ..

Align seg 1/1 to: AAR47575 from: 1 to: 650

```

10 ATTTATTTATTTTACATGTTTCTCTCAGTACTT..... 48
   |||||.....|
   8 IleuTrpLeuPheSerLeuAlaLeuAlaSerAlaIleAsnIleAsnAs 24
49 .TTTACATTGCT.....GTATCGCAGATAAAATTC 79
   ::|||::
24 PileThrPheSerAsnLeuGluIleThrProLeuThrAlaAsnLysGlnP 41
80 CCGGA.....GATGAAGACATAATTAATTTT 108
   ||| |||||::
41 RoAspGlnGlyTrpThrAlaThrPheAspPheSerIleAlaAspLase 57
109 GCGCCCGCTGCAGAACGAA.....TCTCCGCC.....AA 140
   ::|||::|
58 SerIleAspGlnGlyAspGluPheThrLeuSerMetProHisValTyrAr 74
141 ACATTAATATTAATAATACATATTACAGCA..... 171
   :|||:|
74 GileysLeuLeuAsnSerSerGlnThrAlaThrIleSerLeuAlaAsp 91
172 .....TACAGTGAAGTCATCTCTGTAT 195
   ::|||::
91 LyrThrGluAlaPheLysCysTyrValSerGlnGlnAlaIleTyrLeuYr 107
196 GATAGAGAGACTTTTATGT.....TTGCTCTTCACAA 230
   ::|||::|
108 GluAsnThrThrPheThrCysThrAlaGlnAsnAspLeuSerSerTyrAs 124
231 TACACTTAATGAGCATGTCCACACAGATGAGATCCATGAGTTCATCGG 280
   |||||::|
124 nThrIleAspLysSerIleThrPheSerLeuAsnPheSerAspLysGlys 141
281 TCACGGGTGAACAATATTAACATTACAAATTACGAAAAAGAAAGTTTA 330
   ||| ||| ::|||
141 eSerTyrGluTyrGluLeu.....GluAsnAlaLysPhe 152
331 ATAAAAAGAGCGCTACAATTAAGGCTATTAACAATTATGTTCAAAAG 380
   ::||| ::|||
153 PheLysSerGlyProMetLeuValLysLeuGlyAsnGlnMetSerAspVa 169
381 TGTTAACTGC...CCATCGGCGCTAACCTTAACCTCAGCTCATTTTACT 427
   ||||| ||||| ::|||
169 lValAsnPheAspProAlaAlaPheThrGluAsnValPheHisSerGlyA 186
428 GTAATAAAAACGCGCTTCAGGTCAAGCTTATATTA.....TATATT 471
   ::||| ::|||
186 rGserThrGlyTyrGlySerPheGlnSerTyrHisLeuGlyMetTyrCys 202
472 CCTGCTGGCGAAGTAAAAAATTTGCTTTTGCTGATCTGGATGCTAC 521
   ||| ||| ::|||
203 ProAsnGlyTyrPhe.....LeuGlyGly.....Th 211
522 TCTGAAGTTAAGATAAAAAAGACGATATAGTACACTATGGAAGTTCACA 571

```

```

211 RGLuLysIle.....AspTyrIleAspSerSera 220
   |||::|
572 CTATTAATATATCATTTTAATTAAGTAAAGCAAAATATTCAGATA... 618
   ::|||::|
220 sAsnAsnValAspLeu.....AspCysSerSerValGlnValTyr 233
619 .....TGCTTACCTCAGTTCACAAAGAGCAGCC 644
   |||||
234 SerSerAsnAspPheAsnAspTrpThrPheProGlnSerTyrAsnAspTh 250
645 TCGCGTCGATCTTAACCTTGCGTCCAACTGAGGGGCAATATATGGA 694
   ::|||::|
250 rAsnAlaAspVal.....ThrcysPheGlys 259
695 GAAT.....TCTGTGATATGTCCTTTATGATGA..... 726
   ||| ::|||
259 eRAsnLeuTrpLleThrLeuAspGluLysLeuTyrAspGlyGluMetLeu 275
726 ..... 726
276 TrpValAsnAlaLeuGlnSerLeuProAlaAsnValAsnThrIleAspHi 292
727 .....T 727
292 sAlaLeuGluPheGlnTyrThrCysLeuAspThrIleAlaAsnThrTrt 309
728 ATAGTACTAACAGCAGCTCTTTGGAG.....ATPAGATTTACAGATAAC 771
   |||||::|
309 YrAlaThrGlnPheSerThrThrArgGluPheIleValTyrGlnGlyArg 325
772 AAT.....CCTAATCTGATGGGAATTTATCTAAGCAAAATTAAGA 815
   ||| ||| ::|||
326 AsnLeuGlyThrAlaSerAlaLysSerSerPheIleSerThrThrThr 342
816 TGACACAAAGAAAT.....GCATATACCTTGTCA..... 846
   ||| ::|||
342 rAspLeuThrSerIleAsnThrSerAlaTyrSerThrGlySerIleSer 359
847 .....CTTCTC 852
359 hrValGluThrGlyAsnArgThrThrSerGluValIleSerHisValVal 375
853 TTGCGCGGTAAAGTTTAACTCCAAACAATGAGACGTCATTAATATTCG 902
   ::|||::|
376 ThrThrSerThrLysLeuSerProThrAlaThrThrSerLeuThrIleAl 392
903 TGACGACAGCTCTCGCAACAAACTGGAAT..... 933
   |||||
392 aGlnThrSerIleTyrSerThrAspSerAsnIleThrValGlyThrAspI 409
934 .....AGAAAT 939
409 LeuHisThrThrSerGluValIleSerAspValGluThrIleSerArgLys 425
940 ACAGCTGTCACCAAGCCGAAATCAGTGTCCGGGTGTGTGGCCCGG 989
   |||||
426 ThrAlaSerThrValValAlaAlaProThrSerThrThrGlyTrpThrG 442
990 ACGTTTG 996
   | ::|||
442 yAlaMet 444
seq_name: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:AA201833
seq_documentation_block:
ID AAB01833 standard; Protein; 1005 AA.
XX
AC AAB01833:
XX
DT 11-SEP-2000 (first entry)
XX
DE Haemophilus influenzae strain K21 mature HMW2A protein, SEQ ID NO:41.

```



```

806 AATAATATGATGACCAAGAAATTCATATATCTTGTCTCTCTG 855
      |||||
582 snleuthrllelysthrlygluleuthrleuthraspansleuAsnle 598
856 GCGGCT.....AAAGTTTAAGTCCACAAATGGAACGTCATTAA 896
      |||||
599 SerGlyPheAsnLysAlaGluIleThrAlaLysAspAsnSerSerp 615
897 TATTGTCAGCGCTCTCTCTGGAACAAAGTCGATAGATTAACAGCT 946
      |||||
615 eileeglyLysAlaSerSerAspAsnSerAsnAlaLysGlnIleThrPhe 632
947 TCACCATGCCAGAAATCAGTGTCCGCTGTGTGTCGCGCTGACGCTTG 996
      |||||
632 splysValLysAspSerLysIleSer.....AlaGlyAsnHisAsnVal 646
997 CAATTGGATGCAAAAGTGCAGAAATCCCGAGGCT..... 1029
      |||||
647 ThrLeuAsnSerLysValGluThrSerAsnSerAspGlySerThrGly 663
1030 .GGACAATATATGGTAAATATTAATGTTACTTCCACACCAAGATCAAA 1078
      |||||
663 nglySerAspAsnAsnIleGlyLeuThrIleSerAlaLysAspValT 680
1079 CACTC 1083
      |||||
680 hrVal 681

```

seq_name: /SIDSL/gcgcdata/hoid-geneseq/geneseq-emb1/AA1993.DAT:AA41732

seq_documentation_block:

ID AAR41732 standard; Protein: 1529 AA.

AC AAR41732;

DT 26-APR-1994 (first entry)

DE High molecular weight protein 4 (HMP4).

KW HMP: high molecular weight protein; virus; vaccine; influenza;

KX epitope; immunity; haemophilus influenzae.

OS Haemophilus influenzae.

PN W09319090-A.

PD 30-SEP-1993.

PF 16-MAR-1993; 93WO-US02166.

PR 16-MAR-1992; 92GB-0005704.

PA (BARE/) BARENKAMP S J.

PI (INRM) INSERM INST NAT SANTE & RECH MEDICALE.

PI Barenkamp SJ;

DR WPI: 1993-320683/40.

DR N-PSDB: AAO49511.

PT High molecular weight surface proteins - of non-typeable
haemophilus which exhibit immunogenic properties

PS Claim 6; Figure 10; 100pp; English.

CC The isolation and purification of the high molecular weight protein
enables the identification of the major protective epitopes of the
protein by conventional epitope mapping. These epitopes can then be
synthesised using standard techniques and incorporated into fully
synthetic or recombinant vaccines.

Sequence 1529 AA;

alignment_scores:

Quality: 110.50 Length: 348
Ratio: 0.594 Gaps: 18
Percent Similarity: 53.448 Percent Identity: 20.690

alignment_block:

us-09-839-894-9 x AAR41732

Align seg 1/1 to: AAR41732 from: 1 to: 1529

```

142 CATATATATATTAATTAACATATTTACAGATACAGTGAAGTCATCT 191
      |||||
744 TyrAsnGluTyrSerLysHis.....AlaIleAsnSerSerHisAsn 758
192 GTATGATAGATGACTTTTATGTTTGCTCTCTCAATACACTTAAG 241
      |
758 u.....ThrIleLeuGlyAsnValThrLeuGlyG 769
242 GACCATGTCCAAACAGTGAAGATCTAGAGTCA..... 276
      ||
769 Ly.....GluAsnSerSerSerSerIleThrGlyAsnIle 780
277 TCGTCAGCGGTGAACAAATATATACATTACATTT..... 312
781 AsnIleThrAsnLysAlaAsnValThrLeuGlnAlaAspThrSerAsn 797
313 .....ACGGAAGAAAGAGTTTAATTAAGAGAGCTACAAATTA 352
      |||||
797 rAsnThrGlyLeuLysLysArgThrLeuThrLeuGlyAsnIleSerValG 814
353 AAGCTATATAACAAATATGTTTCAAAAGTGTAACTGCCATCCGCGCTA 402
      |||||
814 LucIlyAsnLeuSerLeuThrGlyAlaAsnAlaAsnIleValGlyAsnLeu 830
403 ACACCTTAAGTCACTCATTTTAACTGTAATTAAGACGCGCTTCAGTGC 452
      |||||
831 SerIle.....AlaGluAspSerThrPheLysGlyAlaAlaSer..... 843
453 AAGTTTATATTTTATATATTCCTCTGCGCAACTAAGAAATTTCCCTTG 502
      |||||
844 .....AspAsnLeuAsnIleThr 849
503 GTGCTATCTGG.....GATGCTACTCTGTAAGTTAAGATAAGACGA 546
      |||||
849 hrcGlyThrPheThrAsnAsnGlyThrAlaAsnIleAsnIleLysGlnGly 865
547 TATAGTACAGACTATGGAACCTTACCT.....ATAATATCAC 584
      |||||
866 ValValLysLeuGlnGlyAspIleIleAsnLysGlyGlyLeuAsnIleThr 882
585 TATTAATTTACT..... 597
882 rThrAsnAlaSerGlyThrGlnLysThrIleIleAsnGlyAsnIleThrA 899
598 ..GATAAGGAATATTCAGATATGTTACCTCAGTTCACAAAGTCAGCT 645
      |||||
899 snGluLysGlyAspLeuAsnIle.....LysAsnIleLysAlaAspAla 913
646 CCGGTCGATCTT.....AACTTGCCTCAACTGTGGGGGCAATATAT 689
      |||||
914 GluIleGlnIleGlyLysAsnIleSerGlnLysGlnGlyAsnLeuThrI 930
690 TGAAGAAATTCGTGATATATGCTTTATATGATATAGATACATACAA 739
      |||||
930 eSerSerAspLysValAsnIle.....ThrAsnG 940
740 GCAGCTCTTGGAGATACATTCAG.....GATAACAATCTCT 777
      |||||
940 InleThrIleLysAlaGlyValGlnGlyGlyArgSerAspSerSerGlu 956
778 AATCTGATGGAAATTTATCTAAGCAAAATTAATGATGACACCAAGA 827
      |||||

```


1002 eSerSerAspLysValAsnIle.....ThrAsnG 1012
740 GCAGCCTTTGGAGATAGATTTCAG.....GATACATTCCT 777
:: ::::: ::::: GATACATTCCT 777
1012 InlethrlleLysAlaGlyValGluGlyGlyArgSerAspSerGlu 1028
778 AATCTGATGGGAAATTTATCTAAGAAATTAATGATGACACCAAGA 827
::: ::::: ::::: 11111111
1029 AlaGluAsnAlaAsnLeuThrIleGln.....ThrLysG 1040
828 AATGCAATGATCTTCTCAGCTTCTTGGCGGT.....AAAGTT 868
::: ::::: ::::: 11111111
1040 uLeuLysLeuAlaGlyAspLeuAsnIleSerGlyPheAsnLysAlaGlu 1057
869 TAACTCCACAATAGCAGCTCATTAATGCTGACGAGCTTCCTG 918
::: ||| ::::: ::::: 11111111
1057 IeThrAlaLysAsnGlySerAspLeuThrIleGlyAsnAlaSerGlyGly 1073
919 GAAACAACTGGAAATGAAATTAACAGCTGCACCATGCCAGAAATCAGT 968
::: ::::: ::::: 11111111
1074 AsnAlaAspAlaLysLys.....ValThrPheAspLysValLysAs 1087
969 TCCGGTGTGTGTGGCTTGA...CGTTGCATTTGATGCAAAAGTG 1015
::: ||| ::::: 11111111
1087 pSerLysIleSerThrAspGlyHisAsnValThrLeuAsnSerGluVal 1104
1016 AAAATCCGAGGCTGGCAATATATGCTAATATTAATGTTACT 1059
::: ::::: ::::: 11111111
1104 yThrSerAsnGlySerSerAsnAlaGlyAsnAspAsnSerThr 1118
seq_name: /SIDS1/gcgdata/hold-geneseq/genesep-emb1/AA2000.DAT:AAAB15945
seq_documentation_block:
ID AAB15945 standard; Protein: 2383 AA.
XX AAB15945:
AC
XX
DT 05-OCT-2000 (first entry)
XX
DE E. coli proliferation associated protein sequence SEQ ID NO:302.
XX
KM Escherichia coli; E. coli; proliferation; inhibition; screening;
XX antimicrobial; bacterial growth; antisense therapy; antibacterial.
XX
OS Escherichia coli.
XX
FN WO200044906-A2.
XX
PD 03-AUG-2000.
XX
PE 27-JAN-2000; 2000WO-US02200.
XX
PR 27-JAN-1999; 99US-0117405.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
PI zyskind J, Ohlsen KL, Trawick J, Forsyth RA, Froelich JM, Carr GJ,
PI Yamamoto RT, Xu HH;
XX
DR WPI: 2000-514822/46.
DR N-PSDB: AAA65950.
XX
XX Novel polynucleotides and polypeptides associated with microorganism
PT proliferation, used to identify inhibitors of bacterial growth and
PT proliferation, for use in antisense therapy -
XX
PS Claim 11; Page 224-229; 316pp; English.
XX
CC AAA65809 to AAA65889 and AAA66058 to AAA66138 represent nucleotide
CC sequences derived from Escherichia coli which inhibit E. coli
CC proliferation. AAA65890 to AAA66055 and AAB13886 to AAB16040 represent
CC nucleotide and protein sequences associated with E. coli proliferation.
CC AAA66056 and AAA66057 represent primers used for sequencing E. coli

CC proliferation inhibiting nucleotide inserts in an example from the
CC present invention. Methods from the present invention can be used to
CC identify a proliferation-required gene in a microorganism, by contacting
CC a microorganism with a proliferation-required gene activity inhibitory
CC nucleic acid identified in another organism, and determining if
CC inhibition occurs in the second microorganism. The nucleic acid sequences
CC identified as being required for bacterial growth and proliferation, can
CC be used for antisense therapy for killing bacteria.
XX
SQ Sequence 2383 AA:
alignment_scores:
Quality: 109.00 Length: 390
Ratio: 0.565 Gaps: 15
Percent Similarity: 49.487 Percent Identity: 20.000
alignment_block:
US-09-839-894-9 x AAB15945
Align seg 1/1 to: AAB15945 from: 1 to: 2383
37 TCTTCAGTACTTTTACATTTGCTGATCGGACGATAAATTCGCGAGA 86
||| ::::: ||| ::::: 111
762 SerAlaLysIleAlaThrLeuSerAlaSerAsnAsnGlyValLeuAlaAs 778
87 TGAACCATTAATATATTTTGGCCCGGCGTGACAGAGATCTTCC 136
::: ::::: ::::: 11111111
778 nGluAsnAlaAlaAsnThrValSerValAsnValAlaAspGluLys 794
137 CCAACATATATATTAATTAACCATATTACACATAC.....AGT 177
||| ::::: ||| ::::: 111
795AsnProIleAsnAspHisThrValThrPheAlaValLeuSer 808
178 GAAAGCATACCTCTGTATGATGATGATGACTTTTATGTTGCTTCTCA 227
||| ||| ::::: ::::: 111
809 GlySerAlaThrSerPheAsnAsnGlnAsn.....ThrAlaLys 821
228 CATTAACCTTAATGAGCATGTCACACGAGTGAGATCTAGCATTCAT 277
::: ::::: ||| ::::: 11111111
821 sThrAspValAsnGlyLeuAla...ThrPheAspLeuLysSerLysG 837
278 CGGTCAAGCGGTGAACCAATATTAACATTAACATTAACGAAAGAGT 327
::: ::::: ::::: 11111111
837 InGluAspAsnThrValGluValThrLeuGluAsnGlyValLysGlnThr 853
328 TTAATA..... 333
||| ||| ::::: 111
854 LeuIleValSerPheValGlyAspSerSerThrAlaGlnValAspLeuG 870
334AAAGAGCTACCAATTAAGGCTTAACAATTA...TTGT 373
870 nLysSerLysAsnGluValValAlaAspGlyAsnAspSerValThrMet 887
374 TCAAAGTGTAACTCCATCCGCGCTTAACACTTAACGCTCATTTT 423
::: ||| ::::: ||| ::::: 11111111
887 hrLalThrValAlaArgAspAlaLysGlyAsnLeuLeuAsnspValMetVal 903
424 AACTGTAATTAACAGCGGCTTCAGGTGCAAGTTATTAATTAATATTC 473
::: ||| ||| ::::: 11111111
904 ThrPheAsnValAsnSerAlaGluAlaLysLeuSerGlnThrGluValAs 920
474 TCGTGGCGAACTAAATTTGCTTTGTGTATCTGGAGTCATAC 523
::: ::::: 11111111
920 nSerHisAsp.....GlyIleAlaThrAlaThrL 930
524 TGAAGTTAAGAGTAACAGACGATATAGACACCTATGCACTTACACT 573
||| ||| ::::: 111
930 eurhr.....SerLeuLysAsnGlyAspLysPyrArg 939
574 ATTAATATCATCTTAATTAATTAAGGAAATATTCAGATATGCTT 623
::: ::::: 11111111
940 ValThrAlaSerValSerSerGlySerGlnAlaAsnGlnValAsnph 956


```

334 ILeysSerGlnAsnPhenAlaSerGluGlySer..... 345
300 AACATTCAATTTACGAAAAAGAGCTTAATATAAGAGAGCTACAAA 349
346 SerLeuAArgPhe.....LysS 351
350 TTAAGGCTATTAACATTTATTTGTTCAAAAGTGTAACTGCCATCCGGC 399
351 ergGluGlySerThrArgThrAlaPhe.....ThrIleGluSerAsp 364
400 CTACACTTAACCTACGCTCATTTTAAGTAAATATAACGCGCTGAG 449
365 LeuThrLeuAsnAlaThrGlyAlaSerLeuAsnGlnValAlaG1 381
450 TGCAGTTTATATTTATATATCTGCTGGCGAATAAAATTTGCCCT 499
381 yLeAspGlyAsnLeuGlnLysSerLeuValAlaAsnLysAsnIleThrP 398
500 TTGGTGATCTGGGATGCTACTCTGAAGTTAAGAGTAAAAAGCGATAT 549
398 heGluGly.....GlyAsnIleThrLeuAlaAlaAspLysPro 411
550 AGTGAGACTATGAGACTTACACTATA.....AATACACTAT 587
412 IleGluIleLysGlyAsnIleThrValLysGluGlyAlaAsnValIle 428
588 TAAATTACGTATGAAGGAAATATTCAGATATGTTACTCTGATTCAAA 637
428 uArgSerAlaAsnTyrcIyAsnAspLysSerAlaLeu...SerIleArg 444
638 GTGACGCTCCGTCGATCTTAACCTGCGTCCACATGCTGGGGGACATAT 687
444 LysAsnValIThrAsnLysGlyAsnLeuThrValThrGlySerAlaIleAsn 460
688 ATTGAGAGAAAT.....TCTGTGATATGTGCTTTATGA 722
461 IleGluLysAsnLeuThrValGluGlySerAlaLysPheLeuAlaAsnPr 477
723 TGCATATAGTACTAACAGACGCTTTTGAGATTAAGATTCAGGATACA 772
477 AsnTyrcSerPheAsnValSerGlyLeu.....PheAspSng 490
773 ATCTTAATCTGAT.....CGGAATTTTATCTAAGAAA 807
490 InGlyLysSerAsnIleSerIleAlaLysGlyAlaHisPheLysAsp 506
808 ATAATGATGACACCAAGAATGATATATCTTGTCA..... 846
507 IleAsn...AsnThrLysSerLeuAsnIleThrThrAsnSerAspSerAl 522
847 .....CTTCTCTGGCGGTAAAGTTTAACCTCAACAATGGAAGCT 889
522 aTyrcArgThrIleIleGluGly...AsnIleThrAsnSerAsnLys...A 537
890 CATTAATATATGCTGAC 906
537 sPLeuAsnIleThrAsp 542

```

seq_name: /sids1/gcgdata/hold-geneseq/geneseqp-emb1/AA1996.DAT:AA85290

seq_documentation_block:

ID AA85290 standard; Protein: 671 AA.

XX AA85290;

XX 02-APR-1996 (first entry)

XX Streptococcus faecalis autolysin.

XX Lysin; autolysis; culture; lactic acid bacteria; fermentation;

XX cheese; foodstuffs; induction.

```

OS Streptococcus faecalis.
XX
XX WO9531561-A1.
XX
XX 23-NOV-1995.
XX
XX 12-MAY-1995; 95MO-NL00170.
XX
XX 12-MAY-1994; 94EP-0201353.
XX
XX (UNIL ) QUEST INT BV.
XX
XX Buist G, Kok J, Ledebor AM, Venema G;
XX
XX WPI; 1996-010946/01.
XX
XX Lysis of a culture of lactic acid bacteria in, e.g. cheese
XX production - by in situ prodn. of an auto-lysin, regulated by an
XX inducible promoter.
XX
XX Disclosure; Page 66-69; 103pp; English.
XX
XX In situ production of a homologous autolysin or a heterologous
XX autolysin from a food grade gram positive bacteria, can be used in
XX a process for the lysis of a culture of lactic acid bacteria. The
XX process can be used in the manufacture of products containing
XX cultures of lactic acid bacteria e.g. cheese, where the culture is
XX lysed following the completion of fermentation. The enhanced
XX fermentation of the autolysin is performed some hours after the
XX lysis does not need to be isolated or encapsulated. The time of
XX lysis can be precisely controlled. This is the Streptococcus
XX faecalis autolysin.
XX
XX Sequence 671 AA:

```

alignment_scores:

Quality:	106.00	Length:	309
Ratio:	0.716	Gaps:	17
Percent Similarity:	47.896	Percent Identity:	23.625

alignment_block:

US-09-839-894-9 x AA85290 ..

Align seg 1/1 to: AA85290 from: 1 to: 671

```

121 AGSAGCAATGCTCCCAACATATATA...TTAATACATATTAC 167
168 AGCATACAGTGAAGATCACTCTGTATGATGAGTACATTTTATAGT 217
309 ArgTyrcAlaThrAspProSerTyrcAsnAlaLysLeuAsnValIleThr 325
325 rAlaTyrc.....AsnLeuThrGlnTyrcAspThrProSerSerGlyAla 340
218 TGCTCTCACATATACACTTAATGAGCATGCTCCACCGATGAGATCT 267
340 snThrGlyGlyGlyThrValAsnProGlyThrGlyGlySerAsnGln 356
268 AGCAGTTCATCGGTC.....AGCGTGAACA.....AA 296
357 SerGlyThrAsnThrTyrcTyrcThrValLysSerGlyAspThrLeuAsnLys 373
297 TATACATTTACATTT.....ACGAAAAAAGAGTTTA... 330
373 sIleAlaIleGlnTyrcIyAlaSerValAlaAsnLeuArgSerTyrcSng 390
331 ..ATAAAAGAGAGCTTCAATTAAGCTATTAACAATTAATTTGTTCAA 378
390 LysLeuSerGlyAspLeuIlePheValGlyGlnLysLeuIleValLysLys 406
379 AGGTTAACGCGCCATCCGCGCTTACACTTAACGCTCATTTTACTG 428

```

```

407 GlyAla.....SerGlyAsnThrGlyGlySerGlyAsnGlyGlySer 420
429 TAAATAAAACGGCGCTCAGGTGCAAGTTATATATATATATTCCTG 478
      |||:||||| |||:||||| |||:||||| |||:|||||
420 rAsnAsnAsn...GlnSerGlyThrAsnThrTyTyrThrValIysSerG 436
      |||:||||| |||:||||| |||:||||| |||:|||||
479 GCGAACTAAAAATTGCGCTTTGGTGGATCTGGGATGCTACTGAG 528
      |||:||||| |||:||||| |||:||||| |||:|||||
436 LysAspThrLeuAsn..... 440
529 TTAAGAGTAAAAAGACATATAGTACACTGATGAACTTACTACTAAA 578
      |||:||||| |||:||||| |||:||||| |||:|||||
441 .....LysIleAlaIleAlaGlnTyGly..... 447
579 TATCACTATTAAATTAAGTAAAGAAATATTCACATATGCTTACTCTG 628
      |||:||||| |||:||||| |||:||||| |||:|||||
448 ValThrVal.....AlaAsnLeuArgSerTrp..... 456
629 AGTTCAAAAGTACGCTCGCGTCACTTAACTTG..... 663
457 .....AsnGlyIleSerGlyAspLeuIlePheValGlyGlnIleu 470
664 .....CCTCCAACTGGTGGGCGCACATATATTGGAAGAAATCTGT 704
      |||:||||| |||:||||| |||:||||| |||:|||||
471 IleValIysLysGlyThrSerGlyAsnThr..... 480
705 TGATATGCTGTTTATGATAGTATAGTAAACAGCGCTTTGGAGA 754
      |||:||||| |||:||||| |||:||||| |||:|||||
481 .....GlyGlySerSerAsnGlyIleSer..... 488
755 TAAGATTTCAGATACCAATCTAAATCTGATGGAAATTTATCTAAG 804
      |||:||||| |||:||||| |||:||||| |||:|||||
489 .....AsnAsnAsnGlnSerGlyThrAsnThrTyTyrThrIle 501
805 AAATAATATGATGACACCAAGAAATTTGCA.....TATCTTTGCA.. 846
      |||:||||| |||:||||| |||:||||| |||:|||||
502 LysSerGlyAspThrLeuAsnLysIleAlaGlnTyGlyValSerVa 518
847 .....CTTCTCTTGGCGG 859
518 IAlaAsnLeuArgSerTrpAsnGlyIleSerGlyAspLeuIlePheValG 535
860 GTAATACTTAACCTCAACAATGGAAGTCAATTAATTTGTCGACGCA 909
      |||:||||| |||:||||| |||:||||| |||:|||||
535 LysGlnIleIleValIysLysGlyThrSerGlyAsnThrGlyGlySer 551
910 GCTTCTCGAACAACCTGGAATAGA 936
      |||:||||| |||:||||| |||:||||| |||:|||||
552 SerAsnGlyLysSerAsnAsnAsnGln 560

```

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:AA01830

seq_documentation_block:

ID AAB01830 standard; Protein; 1222 AA.

AC AAB01830;

DT 11-SEP-2000 (first entry)

DE H. influenzae strain KI mature full-length HMW1A protein, SEQ ID NO:37.

KW Mature HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;

KW non-typable Haemophilus influenzae; NTHI; non-encapsulated;

KW recombinant production; Escherichia coli; antibacterial; vaccine;

KW human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;

KW detection; diagnosis.

OS Haemophilus influenzae strain KI.

PH Key Location/Qualifiers

FT Misc-difference 307

FT /note= "Encoded by GG"

XX

PN WO200020609-A2.

PD 13-APR-2000.

PF 07-OCT-1999; 99WO-CA00938.

PR 07-OCT-1998; 98US-0167568.

PR 08-DEC-1998; 98US-0206942.

PA (CONN-) CONNAUGHT LAB LTD.

PI Loosmore SM, Yang Y, Klein MH;

DR WPI: 2000-303789/26.

DR N-PSDB: AAA52180.

PT Nucleic acid molecule for producing recombinant high molecular weight

PT proteins of Haemophilus which are used as a vaccine to provide

PS protection against Haemophilus induced diseases in humans -

CC Claim 8; Fig 20A-R; 307pp; English.

CC The invention relates to the recombinant production of Haemophilus

CC influenzae high molecular weight (HMW) proteins in Escherichia coli. The

CC expression construct used to effect recombinant expression comprises a

CC promoter functional in E. coli (e.g., the T7 promoter) operably linked

CC to a modified hmwABC operon from a non-typable (non-encapsulated) H.

CC influenzae (NTHI). Most HMW-expressing NTHI strains contain two hmw gene

CC clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,

CC hmwB and hmwC genes. The hmwA genes encode the structural HMWA proteins

CC and the hmwB and hmwC genes encode accessory proteins which are

CC responsible for post-translational processing and secretion of the HMWA

CC proteins. The modified hmwABC operon used in the expression construct of

CC the invention contains an A gene modified such that it encodes only the

CC mature HMWA. The invention also discloses hmwA genes (AA52175-A52198)

CC and HMWA proteins (AAB01824-B01849) from the non-typable H. influenzae

alignment_scores:

Quality: 106.00

Ratio: 0.482

Percent Similarity: 48.780

Percent Identity: 20.177

alignment_block:

US-09-839-894-9 x AAB01830

Align seg 1/1 to: AAB01830 from: 1 to: 1222

31 TTTTTCCTTCAGTACTTTTACATTTGCTATCGGACAGAT..... 72

265 PheAsnAsnAspThrValPheAsnIleAlaIleAspSerValAlaAsnPh 281

73AAATTCGCGAGATGAAGCATAACTAATATTGTCGCCCGC 115

281 eAsnIleLysProProIleValAspLysValThrAsn..... 293

116 GTGACAGGAAGATCTTCCCGAACAATATATTA..... 150

294GlyAsnIleThrLeuPheLysGlyAsnIleSerValLeuGlyGly 308

```

151 .....TTAATAACAAATTACAGCATACAGTGAAGTCACTACTCTCTA 194
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
309 GlyAspValAsnPheHisPheAsnLaserSerAsnTyrGlnThrTyr 325
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
195 TGATAGCATGACTTTTATGTGTCTCTCACAAT..... 231
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
325 rGly.....ValIleIleGluSerGlnAsnPheSerAlaSerG 338
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
231 ..... 231
338 LysGlySerSerLeuLysPheLysSerGluGlySerThrHisAlaAlaPhe 354
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
232 .....ACACT 236
355 ThrIleLysAsnAspLeuIleLeuAsnAlaThrGlyLysAsnIleSerLe 371
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
237 TAATGAGCATGTCACACCGATGAGAAATCTAGAGTTCATCGGTACGCG 286
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
371 uAsnGlnValAlaGlyIleAspSerAsnLeuLysSerLeuIleAla 388
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
287 GTGAACA.....AATATACATTACAATTACGGA 318
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
388 snLysAsnIleThrPheGluGlyLysAsnIleThrLeuAlaAlaAspLys 404
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
319 AAAAGAAGTTAATATAAAGAGACTACAAATTAAGGCTTAACAATT 368
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
405 LysProIleGluIleLysGlyLysAsnIleThrValLysGluGlyAlaAsnVa 421
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
369 ATTGTCAAAAGTGTAACTGC..... 390
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
421 LThrLeuArgSerAlaAsnTyrGlyAsnAspLysSerAlaLeuSerIleA 438
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
391 .....CCATCGCGCTACACCTTAACCTACGCTCATTTT 423
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
438 rGlyAsnValThrAsnLysGlyAsnLeuThrValThrGlySerAlaIle 454
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
424 AACGTATATAAANC..GCGGCTTCAGCTGCACAGTTATATTATATAT 470
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
455 AsnIleGluLysAsnLeuThrValGlnLysSerAlaLysPheLeuAlaAs 471
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
471 TCCTGCTCGCACTAAAAAATTGCTCTTT.....GGTGATATCGGG 514
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
471 nPro.....AsnTyrSerPheAsnValSerGlyLeuPheA 483
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
515 ATGCTACTCTGAGCTAAGCTAATAA.....AGACGATAT 549
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
483 sPasnGlnGlyLysSerAsnIleSerIleAlaLysGlyAlaIlePhe 499
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
550 AGTGAGACTTGAAGCTTACACTATAATATCATATAATTAATTAATGA 599
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
500 LysAspIleGluAsnThrGlySerLeuAsnIleThrThrLysSerAspSe 516
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
600 T.....AAGGAAATATTCAGATATGTTACTCTAGT 631
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
516 rAsnHisHisThrIleIleLysGlyAsnIle..... 526
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
632 TCAAAAGTACGCTCGCTCGATCTTAACCTGCTCACTCACTGCTGGGCG 681
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
527 .....ThrAsnArgLysGlyAspLeuAsnIleThrAsnAsnGlyAspAsn 541
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
682 ACA.....TATATTGGAAGAATTCGTGTATATGTCCTTTTATGATGG 725
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
542 ThrGlnIleGlnIleGlyGlyAsn.....IleSerGlnLysGluG 555
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
726 ATATAGTACTAAGACAGCTCT.....TTGAGA 754
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
555 yAsnLeuThrIleSerSerAspLysValAsnIleThrGlnArgIleHri 572
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
755 TAAGATTTACAGATACAAATCTTAATCTGAT...GGGAAATTTTATCTA 801
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
572 lLysAlaIleGlyValAsnGlyAspAsnSerAspSerAsnGlnAlaIleThrSer 588

```

```

802 AGGAAATATATGATGACACCAAGAAATTCGATATFACTTTGTCACTCT 851
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
589 AlaAsnLeuThrIleLysThrLysGluLeuLysLeuThrAsnAspLeuAs 605
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
852 CTGGCGGGT.....AAAAGTTTAACTCACAAACAATGGAAGCTCAT 892
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
605 nIleSerLysPheAsnLysAlaGlnIleThrAlaLysAspAsnSerAsnL 622
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
893 TAAATATTGCTGACCGCAGCTTCTCTG...GAACAACAATGCAATGAA 939
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
622 eutnIleGlyAspAsnSerAspAlaGlyAsnThrAspAlaLysLysVal 638
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
940 ACAGCTGCACCATGCCGAAATCAGTGTCCGGGTGCTGTGGCCCG 989
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
639 ThrPheSerAsnValLysAspSerLysIleSerAlaSer.....AspHi 653
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
990 ACGTTTGCAATTGGATGCAAAAGTGGA.....AATCCGAGGCTG 1030
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
653 sAsnValThrLeuAsnSerLysValGlnThrSerGlyAspThrAspSerT 670
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1031 GACAAATATATGGTATATATATGTTACTTTCACACCAAGTAGTCAACA 1080
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
670 hrgLyspGlyGlyAsnAsnAsnThrGlyLeuThrIleThrAlaLysAsn 686
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1081 CTC 1083
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
687 Val 687
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA2000.DAT:AA801828
seq_documentation_block:
ID   AAB01828 standard; protein; 1228 AA.
AC   AAB01828;
DT   11-SEP-2000 (first entry)
DE   Haemophilus influenzae strain KI full-length HMW1A protein. SEQ ID NO:34.
KW   HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KW   non-typeable Haemophilus influenzae; NMHI; non-encapsulated;
KW   recombinant production; Escherichia coli; antibacterial; vaccine;
KW   human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KW   detection; diagnosis.
OS   Haemophilus influenzae strain KI.
OX   OS
FH   Key Location/Qualifiers
FT   Misc-difference 313 /note="Encoded by GG"
FT   WT0200020609-A2.
PD   13-APR-2000.
PE   07-OCT-1999; 99MO-CA00938.
PR   07-OCT-1998; 98US-0167568.
PR   08-DEC-1998; 98US-0206942.
PA   (CONN-) CONNAUGHT LAB LTD.
XX   Loosmore SM, Yang Y, Klein MH;
XX   WPI; 2000-303789/26.
XX   DR N-PSDB; AAA52179.
XX   PT Nucleic acid molecule for producing recombinant high molecular weight
XX   PT proteins of Haemophilus which are used as a vaccine to provide
XX   PT protection against Haemophilus induced diseases in humans -
XX   Claim 12; Fig 20A-R; 307pp; English.
XX

```


676 hrgluaspGlyAsnAsnAsnThrGlyLeuThrIleThrAlaLysAsn 692
1081 CTC 1083
693 Val 693

seq_name: /SID1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT:AA41731

seq_documentation_block:

ID: AA41731 standard; Protein; 1338 AA.

AA41731;

26-APR-1994 (first entry)

High molecular weight protein 3 (HWP3).

HMW; high molecular weight protein; virus; vaccine; influenza;
epitope; immunity; haemophilus influenzae.

Haemophilus influenzae.

WO9319090-A.

30-SEP-1993.

16-MAR-1993; 93WO-0502166.

16-MAR-1992; 92GB-0005704.

(BARE/) BARENKAMP S J.
(INRM) INSERM INST NAT SANTE & RECH MEDICALE.

Barenkamp SJ;

WPI: 1993-320683/40.

N-PSDB: AA049510.

High molecular weight surface proteins - of non-typeable
haemophilus which exhibit immunogenic properties

Claim 5; Figure 10; 100pp; English.

The isolation and purification of the high molecular weight protein
enables the identification of the major protective epitopes of the
CC protein by conventional epitope mapping. These epitopes can then be
CC synthesised using standard techniques and incorporated into fully
CC synthetic or recombinant vaccines.

Sequence 1338 AA;

alignment_scores:

Quality: 101.00 Length: 418

Ratio: 0.529 Gaps: 18

Percent Similarity: 45.694 Percent Identity: 21.531

alignment_block:

US-09-839-894-9 x AA41731 ..

Align seg 1/1 to: AA41731 from: 1 to: 1338

61 GATCGGCGACAGATAATTCGCCGAGATGAAGCATAACTAATATTTTGG 110
||||: |||||:
817 ValThrPheAspLysVal..... 822
111 CCCGCGTACAGACGATCTTCCCAACAAATATATA...TTAATA 157
||||: |||||:
823 ...LysAspSerIleSerThrAspGlyHisAsnValThrLeuAsn 838
158 ACATATAT...ACAGCATACAGTGAAGATCATACTGTATGATAGATG 204
||||: ||||: ||||: ||||:
838 ergluValLysThrSerAsnGlySerSerAsnAlaGlyAsnAspAsnSer 854

205 ACTTTTATCTTGTCT..... 222
||| |||
855 ThrGlyLeuThrIleSerAlaLysAspValThrValAsnAsnAsnValThr 871
223 ..TCTCACAATACACTTAAT.....GGAGCATGCTCAACCA 256
||||: ||||: ||||: ||||:
871 rSerHisLysThrIleAsnIleSerAlaAlaGlyAsnValThrThrL 888
257 GTGAGAAATCTGACGATTCAGTGGTCAAGCGGTGAACAAATATACATTA 306
||||: |||: ||||: ||||: ||||: ||||: ||||: ||||:
888 ysgluGlyThrThrIleAsnAlaThrThrGlySerValGluValThrAla 904
307 CAA.....TTTAC 314
||| |||
905 GluAsnGlyThrIleLysGlyAsnIleThrSerGluAsnValThrValTh 921
315 GGAATAAAGAGTTTAATAAAGAGACTCAAAAT.....A 352
| |||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:
921 rAlaThrGluAsnLeuValThrThrGluAsnAlaValIleAsnAlaThrS 938
353 AAGGCTATTAACAATTTATGTTCAAAAGTGTAACTGCCATCGGCCTA 402
||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:
938 erglyThrValAsnIleSerThrLysThrGlyAspIleLysGlyLysLe 954
403 ACACCTAACCTGAGCTCATTTTAACTGTAATAAAAACGCGCTTCAAGTGC 452
|||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:
955 GluSerThrSerGlyAsnValAsnIle.....ThrAlaSerGlyAs 968
453 AAGTTTA..... 459
||| |||
968 nThrLeuLysValSerAsnIleThrGlyLysAspValThrValThrAla 985
460TATTTAATATTCCTGCGCACTAAATAAATTCCTTT 501
||| ||||: ||||: ||||: ||||: ||||: ||||: ||||:
985 spAlaGlyAlaLeuThrThrThrAlaGlySerThrIleSerAlaThrThr 1001
502 GGTGCTATCTGCGATGCTACTCTCAAGTTAAGACTAAACACATATAG 551
||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:
1002 GlyAsnAlaAsnIleThrLysThrGlyAspIleAsnGlyLysValG 1018
552 TGAGACCATGACCTACACTATAATATATCACTAATTAATTAATCTGATA 601
|||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:
1018 uSerSerSerGlySerValThrLeuValAlaThrGlyAlaThrLeuAlaV 1035
602 AGGGAATATATTCAGATATGTTACCTCAGTTCAAAAGTACGCTGCGCT 651
||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:
1035 alGlyAsnIleSerGlyAsnThrValThrIleThrAlaAspSerGly... 1050
652 GATCTTAACCTTGGCTCCACTGTGGGGGACACATATATTTGAGAAATTC 701
|||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:
1051LysIleuThrSerThrValGlySerThrIleAsnGlyThrAsnSe 1065
702 TGTGATATGTGCTTTATATGATGATATAGTACAAACGACGCTTTGG 751
||| ||||: ||||: ||||: ||||: ||||: ||||: ||||:
1065 rValThrThr.....SerSerGluSerGlyAspIleG 1076
752 AGATAGATTTCAGATTAACAAT.....CTTAATCTGATGGGAAA 792
|||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:
1076 luGlyThrIleSerGlyAsnThrValAsnValThrAlaSerThrGlyAsp 1092
793 TTTTATCTA...AGAAATTAATGATGACACCAAGAAATTCATATAC 839
|||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:
1093 LeuThrIleGlyAsnSerAlaLysValGluValLysAsnGlyAlaAlaTh 1109
840 TTGTGCA.....CTTCTTGGCGGTAAAGTTAA 871
|||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:
1109 rLeuThrAlaGluSerGlyLysLeuThrThrGlnThrGlySerSerIle 1126
872 CTCACAAATGGAAGCTCATTAATATGCTGACGACGCTTCTGTGAA 921
|||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:
1126 hrSerSerAsnGlyGlnThrThrLeuThrAlaLysAspSerSerIleAla 1142

922 ACAAACTGGAAAT.....AGATTACACTGTGCACCAACCCAGAANAATCAG 965
||| ||| ||| :
1143 GlysAlIleAsnAlaAlaAsnValThrLeuAsnTrpThrGlyThrLeuThf 1159
::: :::
966 TGTAT.....CCGGTGTTCGTGGCGTAGACGTTTGCAATG. 1002
::: :::: ||| ||| :::
1159 rTrpTrpGlyAspSerLysIleAsnAlaThrSerGlyThrLeuThrIleA 1176
::: :::
1003GATCCAAGACGGAATAATCCCAGACCGTGA..... 1032
||||||| :::
1176 snAlaLysAspAlaLysLeuAspGlyAlaAlaSerGlyAspArgThrVal 1192
::: :::
1033CAATATATGGTAATATTAATGTTTACTCTCACACCAG 1070
::: ||| :::: ||| ||| |||
1193 ValAsnAlaThrAsnAlaSerGlySerGlyAsnValThrAlaLysThrSe 1209
::: :::
1071 TACT 1074
||| |
1209 rSer 1210


```

195 TGATAGATGACCTTTTATGTTGCTCTCACAAATACCTTAATGAG 244
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
64 rAspArg1LeuAlaPheLeuCythrSerSerSerProValasnGly 81
245 CATCTCCACACGAGTACAGATCTAGACGTTTCATCGTCAGC...GGTGA 291
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
81 lacysProthIle.....GlyThrSerGlyValGlnTyrGlyThr 94
292 ACAAAATATACATTACATTTAGCGAAAAAGAGTTTAAATAAAGAGA 341
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
95 ThrThrIleThrLeuGlnPheThrGluLysArgSerLeuIleLysArg 111
342 GCTACAAATTAAGGCTATAACAAATATCTGTCAAAAGTTTACGCC 391
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
111 nIleasnLeuAlaGlyasnLysProIleTyrGluasnIleSerCys 128
392 CATCCGCG.....CTAACACTTACCTCAGCTCATTTTAACTGAATAA 435
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
128 sPheSerAsnLeuMetValLeuasnSerLysSerTyrPseCysGlyAla 144
436 AACGCG...GCTCAGGTGCAAGTTTATATTATATATCTCTGCGCA 482
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
145 HisGlyasnAlaasnGlyThrLeuLeuasnLeuTyrIleProAlaGly 161
483 ACTAAAAATTTGGCTTTGGTGTATCTGGGATGCTACTCTGAAGTTAA 532
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
161 uIleasnLysLeuProPheGlyGlyIleTyrGluAlaThrLeuIleu 178
533 GAGTAAAAAGACGATATAGTGAGAC.....TATGCAACTTAC 570
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
178 rGleuser...ArgTyrGlyGluValSerSerThrHisTyrGlyAsnTyr 193
571 ACTATAATATCTATTAAATTAAGTGAAGGAAATATATTCAGATAG 620
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
194 ThrValasnIleThrValAsnLeuThrAspLysGlyAsnIleGlnVal 210
621 GTTACCTCAGTCAAAAGTGACGCTCGCTGATCTTACTTGGCTCCA 670
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
210 PheProGlyPheHisSerAsnProArgValAsnLeuasnLeuArgPro 227
671 CTGCTGGGCGCATATATTGAGAGAAATTCGTGATATGCTTTTAT 720
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
227 legIYasnTyrLysTyrSerLysSerasnSerLeuAsnMetCysPheTyr 243
721 GATGATATATGACTTAACAGCAGCTTTTGAGATTAAGATTTCAGATTA 770
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
244 AspGlyTyrSerThrAsnSerAspSerMetValIleLysPheGlnAsp 260
771 CAATCCTAAATCTGATGGAAATTTTATCTAGAGAAATTAATGATGACA 820
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
260 pAsnProThrAsnSerSerGluTyrAsnLeuTyrLysIleGly...GlyT 276
821 CCAAGAAATGATGATATCTTGTACCTCTGCGCGGCTAAAGTTTA 870
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
276 hrgLysLysLeuProTyrAlaValSerLeuLeuMetGlyGluLysIlePhe 292
871 ACTCCAAACAAATGAGACGCTCATTAATAATTCGTGACGAGCTTCTGGA 920
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
293 TyrProValasnGlyInsSerPheThrIleasnSpSerSerValLeuG 309
921 AACCAATCGGAATAGAAATACAGCTGCACCATCCAGAAATCAGTGTG 970
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
309 uThrAsnThrAsnArgValThrAlaValAlaMetProGluValasnValP 326
971 CGGTGTGTGTGGCTGACGCTTGCATTAATGAGTGAAGTGAAGTGA 1020
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
326 rovalleuLysTyrProAlaArgLeuLeuLeuasnAlaAspValAsnAsn 342
1021 CCCGAGGCTGACAAATATATGGTAAATATATGTTACTTTCACACCAAG 1070
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
343 ProGluAlaGlyInsTyrMetGlyAsnIleLysIleThrPheThrProse 359

```

```

1071 TAGTCAACACTC 1083
|||||:|||||:
359 rSerGlnThrLeu 363

```

```
seq_name: /cgn2_6/prodata/2/laa/5B_COMB.pep:US-08-483-101-5
```

```
seq_documentation block:
```

```

; Sequence 5, Application US/08483101
; Patent No. 5932715
; GENERAL INFORMATION:
; APPLICANT: Scott, June R.
; APPLICANT: Froehlich, Barbara
; APPLICANT: Caron, Judy
; TITLE OF INVENTION: CS2 Proteins and Coding Sequences
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee and Winner, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: US
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,101
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33878
; REFERENCE/DOCKET NUMBER: 6-95
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 499-8080
; TELEFAX: (303) 499-8089
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 364 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-483-101-5

```

```
alignment_scores:
```

```

Quality: 927.50 Length: 366
Ratio: 3.198 Gaps: 3
Percent Similarity: 79.235 Percent Identity: 48.087

```

```
alignment_block:
```

```
US-09-839-894-9 x US-08-483-101-5 ..
```

```
Align seg 1/1 to: US-08-483-101-5 from: 1 to: 364
```

```

1 ATGATTAATGATTTATTTATTTTACATTTGTTTCTCTCAGTACTTT 50
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1 LeuLysLysValIlePheValLeuSerMetPheLeuCySerGlnValTyr 17
51 TACATTTCTGTATCGGACGATAAATTCGCGGATGATAAAGCATACTA 100
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
17 rGlyInsThrPheThrAsnValGluAlaGlySerIleasnLysThrG 34
101 ATATTTTGGCCGCGGTGACGAGAAC.....GAATCTTCCCAACAT 144
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
34 luserIleGlyProIleAspArgSerAlaAlaIleSerTyrProAlaHis 50
145 AATATATTAAATTAACATATTACAGCATATGAGTGAAGTCACTACTGTA 194
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
51 TyrIlePheHisGlnHisValAlaGlyTyrAsnLysAspHisSerLeuPh 67

```

```

195 TGATAGGATGACCTTTTATGTTGCTCTGCACAAATACACTTAATGAG 244
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
67 eAspArgMetThrPheLeuCySmetSerSerThrHisAlaSerLysGly 84
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
245 CATGTCCACCAAGAGAGATCTAGCAGTTCATCGGTGACGGGGAACA 294
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
84 lAcysProThrGlyGluAsnSerLysSerSer.....GlnGlyGluThr 98
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
295 AATATACATTTACAAATTTACGAAAAAAGATTAAATAAAAAGAGACT 344
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
99 AsnIleLysLeuIlePheThrGluLysSerLeuAlaArgLysThrIle 115
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
345 ACAATTTAAAGCTATATAACATTATTGTTCAAAAGTTAACTGC.... 390
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
115 uAsnLeuLysGlyTyrLysArgPheLeuTyrGluSerAspArgCysIleH 132
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
391 ....CGATCGGCGCTAACACTTACCTACAGCTCACTTTAACTGAATAA 435
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
132 lSTyValAlaSplysMetAsnLeuAsnSerHisThrValLysCysValGly 148
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
436 AAGCGGCTTCAGGTCGAGTTATATTATATTCCTGCTGCGGCACT 485
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
149 SerPheThrArgGlyValAspPheThrLeuTyrIleProGlnGlyGluI 165
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
486 AAAAAATTTGCTTTGTTGATCTGGAGTCTGACTCTGAAGTTAAGAG 535
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
165 eAspGlyLeuLeuThrGlyIleTyrPgluAlaThrLeuGlnLeuArgY 182
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
536 TAAAAAGACATATAGTACAGCCTATGGAACCTTACACTATAATATGACT 585
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
182 aLysArgHisTyrAspTyrAsnHisGlyThrTyrLysValAsnIleThr 198
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
586 ATTAATATCTGATAGGAAATATTCAGATATGCTTACCTGACGTCACA 635
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
199 ValAspLeuThrAspLysGlyAsnIleGlnValThrProLysPheH 215
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
636 AAGTGCAGCTCGCGTCGATCTTAAGTGCCTCAACTGAGGGGACACAT 685
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
215 sSerAspProArgIleAspLeuAsnLeuArgProGlnGlyAsnGlyS 232
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
686 ATATTTGGAATAAATCTGTGATATGCTTTTATGATGATATATGACT 735
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
232 yrSerGlySerAsnValLeuGlnMetCysLeuTyrAspGlyTyrSerThr 248
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
736 AACAGCAGCTCTTGGAGATAGATTGAGATTAACAATCTTAATCTGA 785
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
249 HisSerGlnSerIleGlnMetArgPheGlnAspSerSerIleThrGlyAs 265
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
786 TGGGAAATTTTATCTAAGAAATTAATGATGACACCAAGAAATTCAT 835
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
265 naSngLutTyAsnLeuIleLysThrGlyIuProLeuLysLysLeuProt 282
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
836 ATACTTTGTACTTCTTGTGGCGGTAAAGTTAACTCCACAAATGGA 885
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
282 yTrLysLeuSerLeuLeuGlnGlyArgGluPheTyrProAsnAsnGly 298
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
886 ACCTCAATTAATATGTGCTGACGAGCTTCTGGAACAACAAGGAATAG 935
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
299 GluAlaPheThrIleAsnAspThrSerSerLeuPheIleAsnThrPsnAr 315
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
936 AATTACAGCTGTGACCATGCCAGAAATCAAGTTCCGGTGTGTGTGGC 985
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
315 gTleLysSerValSerLeuProGlnIleSerIleProValLeuCysTrpP 332
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
986 CTGGACGTTTGCAATTGGATGCAAAAGTGAAAAATCCGAGCTGCACAA 1035
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
332 roAlaAsnLeuThrPheMetSerGluLeuAsnAsnProGluAlaGlyGlu 348
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1036 TATATGGGTATATATTAATTAATCTTACACCAAGTAGTCAACACAGC 1083
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
349 TyrSerGlyIleLeuAsnValThrPheThrProSerSerSerSerLeu 364
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
seq_name: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:US-08-617-697-10

```

```

seq_documentation_block:
: Sequence 10, Application US/08617697
: Patent No. 5977336
: GENERAL INFORMATION:
: APPLICANT: Barenkamp, Stephen J
: TITLE OF INVENTION: High Molecular Weight Surface Proteins
: TITLE OF INVENTION: Of No. 5977336-Typeable Haemophilus
: NUMBER OF SEQUENCES: 11
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Shoemaker and Maltare, Ltd.
: STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
: CITY: Arlington
: STATE: Virginia
: COUNTRY: U.S.A.
: ZIP: 22202-0286
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/617,697
: FILING DATE: 01-APR-1996
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/302,832
: FILING DATE: 05-OCT-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US PCT/US93/02166
: FILING DATE: 16-MAR-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Berkstreser, Jerry W
: REGISTRATION NUMBER: 22,651
: REFERENCE/DOCKET NUMBER: 1038-557
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703) 415-0810
: TELEFAX: (703) 415-0813
: INFORMATION FOR SEQ ID NO: 10:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1600 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: US-08-617-697-10

alignment_scores:
: Quality: 116.50 Length: 343
: Ratio: 0.613 Gaps: 19
: Percent Similarity: 55.394 Percent Identity: 21.574

alignment_block:
US-09-839-894-9 x US-08-617-697-10 ..
Align seq 1/1 to: US-08-617-697-10 from: 1 to: 1600

142 CATTAATATTAATAATCAATATTAACAGCATGAGGAAGCATACTCT 191
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
816 TyrAsnGluTyrSerLysHis.....AlaIleAsnSerSerHisAsnLe 830
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
192 GTATGATAGATGACTTTTATGTTGCTTGTCTCACAATACACTTAAG 241
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
830 u.....ThrIleLeuGlyGlyAsnValThrLeuGlyG 841
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
242 GACCATGTCACACAGAGAGATCTAGCAGTCA..... 276
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
841 Ty.....GluAsnSerSerSerSerIleThrGlyAsnIle 852
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
277 TCGGTACGCGGTGAACAATATTAACATTACAATT..... 312
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
853 AsnIleThrAsnLysAlaAsnValThrLeuGlnAlaAspThrSerAsn 869
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

313 .....ACGMAAAAGAGTTTAAATAAAAGAGCTACAAATTA 352
      |||||
869 rAsnthrclyleuylsYsrgThleuthrleuclYasnIleserValg 886
      |||||
333 AAGCTTAATACAAATTAATGTTCAAAAGTGTAACTGCCATCCGGCTTA 402
      |||||
886 lUglYasnleuSerleuthrclYalaasnAlaasnIleValclYasnleu 902
      |||||
403 ACACCT.....AACTCAGCTCATTTAACTGTAA 431
      |||||
903 SerIlealagluaspSerThrPhelYsglYaluaIaseraSPasnleuAs 919
      |||||
432 T.....AAAACGGCGCTTCAGTGCACAGTTATAT 463
      |||||
919 nIlethrclyThrPhethrAsnAsnglyThrAlaasnIlesnIleYsg 936
      |||||
464 TATATATTCCTGCTGCGCACTAAATAATTCCTTTGGTGTATCTGG 513
      |||||
936 lYValIvalYleuclYaspIleasnAnlys.....GlyGly..... 948
      |||||
514 GATGCTACTGCAAGTTAAGAGTAAAGACATATAGTACAGCTATG 563
      |||||
949 .....leuasnIleThrThrasnAlaSerGlyThrGlnly 960
      |||||
564 AACT...TACATAAATATACATATAATTAACGATAGAGGAATA 610
      |||||
960 sThrIlelleasnclYasnIleThr.....AsnclulYsglYAspL 974
      |||||
611 TTCAGATATGTTTACCTCAGTTCAAAGTGAAGCTGGCGGTGATCT... 657
      |||||
974 euAsnIle.....lysasnIleYsAlaaspAlagluIleGlnleGly 988
      |||||
658 ...AACTGGCTCAACTGGTGGGGCACATATATGGAAGAAATCTGT 704
      |||||
989 glyasnIleserGlnlysglYasnleuthrIleserAspIysVa 1005
      |||||
705 TGATATGCTTTTATATGATATAGTATAGTACAGCAGCTCTTGGAGA 754
      |||||
1005 lAsnIle.....ThrasnIleThrIleYsa 1015
      |||||
755 TAAGATTTTCAG.....GATAACAATCCTTAATCTGATGGAGAA 792
      |||||
1015 lAglyValglYclYarGserAspserSerGluAlaGluasnIlaSn 1031
      |||||
793 TTTTATCTAAGGAAATATATGATGACACCAAGAAATGCAATATCTTT 842
      |||||
1032 leuthrIleGln.....ThrlysgIleuLysleuAlaGl 1043
      |||||
843 GTCACCTTCTTGGCGGT.....AAAGTTTAACTCCACAATG 883
      |||||
1043 yAspIleuasnIleserGlyPheasnlysaIagluIleThrAlaYsaNg 1060
      |||||
884 GAACGCTCAATTAATATGCTGACGAGCTCTCTGGAACAACATGCAAT 933
      |||||
1060 lYserAspIleuthrIleclYasnAlaSerGlyglYAsnAlaaspIaYs 1076
      |||||
934 AGAATTACAGCTGCACCATGCCAGAAATCAAGTTCGGGTGTGTG 983
      |||||
1077 lYs.....ValThrPheaspIlysaSPserlyIleserTh 1090
      |||||
984 GCGTGA...GTTTGAATTGGATGCAAAAGTGAATAATCCCGAGCTG 1030
      |||||
1090 rAspclYHlIsnvalThrleuasnSerGluVallyThrserasnGlyS 1107
      |||||
1031 GACAATATATGCTAATATTAATGTTACT 1059
      |||||
1107 erSerAsnAlaGlyasnAspAsnSerThr 1116

```

seq.name: /cgn2.6/ptodata/2/laa/6A_COMB.pep:US-08-362-525-2

seq_documentation_block:
; Sequence 2, Application US/08362525

```

Patent No. 6027910
GENERAL INFORMATION:
APPLICANT: KITS, FRANCISCUS M.
APPLICANT: SCHREUDER, MARTEN P.
APPLICANT: TOSCHKA, HOLSER Y.
APPLICANT: VERrips, CORNELIS T.
TITLE OF INVENTION: PROCESS FOR IMMOBILIZING ENZYMES TO THE
CELL WALL OF A MICROBIAL CELL BY PRODUCING A FUSION
TITLE OF INVENTION: PROTEIN
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: CUSHMAN DABRY & CUSHMAN, L.L.P.
STREET: 1100 New York Avenue, N.W.
City: Washington
STATE: D.C.
COUNTRY: U.S.A.
Zip: 20005-3918
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/362,525
FILING DATE: 04-JAN-1995
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: EP 92202080.5
FILING DATE: 08-JUL-1992
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: EP 92203899.7
FILING DATE: 14-DEC-1992
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: PCT/EP93/01763
FILING DATE: 07-JUL-1993
ATTORNEY/AGENT INFORMATION:
NAME: KOKULIS, PAUL N.
REGISTRATION NUMBER: 16,773
REFERENCE/DOCKET NUMBER: 213289/T7020(V)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 861-3000
TELEFAX: (202) 822-0944
TELEX: 6714627 CUSH
INFORMATION FOR SEO ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 650 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-362-525-2

alignment_scores:
Quality: 114.00 Length: 469
Ratio: 0.573 Gaps: 23
Percent Similarity: 42.431 Percent Identity: 20.896

alignment_block:
US-09-839-894-9 x US-08-362-525-2 ..
Align seg 1/1 to: US-08-362-525-2 from: 1 to: 650

10 ATTTATTTATTTTACATGTTTCTCTCAGTACTT..... 48
|||||
8 lIleuthrleuPheSerleuAlaIleuAlaIleasnIleasnAs 24
|||||
49 .TTTACATTTGCT.....GTATCGGCGATATAATTC 79
|||||
24 lIleuthrPheSerAsnleuGluIleThrProleuthrAlaasnlysgIlnP 41
|||||
80 CCGGA.....GATGAAGCATACTAATATTTT 108
|||||
41 rAspGlnclYlThrPhrAlaThrPheasnPheSerIleAlaIaspIaSer 57

```

```

772 AAT.....CCTAAATCGATGGCAAAATTTTAACTAAGCAAAATTAATGCA 815
||| |||
326 AsnLeuGlyThrIaSerAlaIylSerPheIleSerThrThrThr 342
||| ::|||
816 TGACACCAAGAAAT.....GCATTAAGTTGTCA..... 846
||| ::||| |||||:::
342 AspLeuThrSerIleAsnThrSerAlaIylSerThrGlySerIleSerT 359
||| ::|||
847 .....CTTCTC 852
359 hrValGluThrGlyAsnArgThrThrSerGluValIleSerHisVal 375
||| ::|||
853 TTGGCGGTAAAGTTTAACTCCCAACAAATGACCATTAATATTC 902
||| ::||| |||||:::
376 ThrThrSerThrIylsLeuSerProThrAlaThrThrSerLeuThrIleAl 392
||| ::|||
903 TGACGCAGCTTCTGTGAAACAACTGAAT..... 933
||| ::|||
392 aglInThrSerIleYrSerThrIylsPheSerAsnIleThrValGlyThrAspI 409
||| ::|||
934 .....ACAATT 939
409 leHsThrThrSerGluValIleSerAspValGluThrIleSerArgGlu 425
||| ::|||
940 ACAGCTGTACACCATGCCAGAACTCAGTTCGCGGTGTGTGGCGCTG 989
||| ::||| |||||
426 ThrIaSerThrValAlaIaIaProThrSerThrThrGlyrPrhrG 442
||| ::|||
990 ACGTTTG 996
||| ::|||
442 yAlaMet 444
||| ::|||

```

```

seq_name: /cgn2_6/prodata/2/1aa/5B_COMB.pep:US-08-728-470-10
seq_documentation_block:
  Sequence 10, Application US/08728470
  Patent No. 5928651
  GENERAL INFORMATION:
  APPLICANT: Barenkamp, Stephen J
  TITLE OF INVENTION: High Molecular Weight Surface Protein
  TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus
  NUMBER OF SEQUENCES: 10
  CORRESPONDENCE ADDRESSES:
  ADDRESSEE: Shemacker and Mattare, Ltd.
  STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
  STREET: Bldg. 1
  CITY: Arlington
  STATE: Virginia
  COUNTRY: U.S.A.
  ZIP: 22202-0286
  COMPUTER READABLE FORM:
  MEDIUM TYPE: Floppy disk
  COMPUTER: IBM PC compatible
  OPERATING SYSTEM: PC-DOS/MS-DOS
  SOFTWARE: PatentIn Release #1.0, Version #1.30
  CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08728,470
  FILING DATE:
  CLASSIFICATION: 424
  PRIOR APPLICATION DATA:
  APPLICATION NUMBER: US 08/302,832
  FILING DATE: 16-MAR-1993
  PRIOR APPLICATION DATA:
  APPLICATION NUMBER: US PCT/US93/02166
  FILING DATE: 16-MAR-1993
  PRIOR APPLICATION DATA:
  APPLICATION NUMBER: GB 9205704.1
  FILING DATE: 16-MAR-1992
  ATTORNEY/AGENT INFORMATION:
  NAME: Berkstresser, Jerry W
  REGISTRATION NUMBER: 22,651
  REFERENCE/DOCKET NUMBER: 1058-633
  TELECOMMUNICATION INFORMATION:

```

314 GUNREGIMILEGILYASMIIESERGINLYSGIUGIYASINLEUINII 930

[illegible]

TELEPHONE: (703) 415-0810
 TELEFAX: (703) 415-0813
 INFORMATION FOR SEQ ID NO: 10:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1529 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 US-08-719-641-10

alignment_scores:

Quality: 111.50 Length: 348
 Ratio: 0.596 Gaps: 18
 Percent Similarity: 53.736 Percent Identity: 20.690

alignment_block:

US-09-839-894-9 x US-08-719-641-10 ..

Align seg 1/1 to: US-08-719-641-10 from: 1 to: 1529

```

142 CATATATATATAATACCATATTACGACATACGACAAAGTCTACTCT 191
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
744 TYSANGLUTYRISERLYSHIS.....AlaIleasnserSerHisasnl 758
192 GTATGATAGACATCTTTTATGTGTCTCTCCACCAATACACTAATG 241
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
758 U.....ThrlleuglyglyAsnValThrlleuglyg 769
242 GACCATGTCACCAACGATGAGATCTTACAGCTCA..... 276
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
769 LY.....GluasnserSerSerIleThrlglyAsnIle 780
277 TCGGTGCGGTGAACAATATACATTAACAATT..... 312
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
781 AsnIleThrlasnlysalasnlValThrlleuglyAlaAspThrserasnl 797
313 .....ACGAAAAAAGATTATATAAAGAGACCTACAATTA 352
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
797 rAsnThrlglyleuLysarThrlleuThrlleuglyAsnIleSerValg 814
353 AAGCGTAAACAATTTTGTCAAAAGTGTAACTGCCATCCGGCCTA 402
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
814 lueIlysnleuSerleuThrlglyAlaAsnIleValglyAsnleu 830
403 ACACCTAACCTCATTTTAACTGTAATRAAAACCGGCTTCAGTGC 452
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
831 SerIle.....AlaGluaspSerThrlPheLysglYlnuIaSer.... 843
453 AAGTTTATTATTATATCTCTGCTGCGAATAAAATTTGCTTTTG 502
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
844 .....AspAsnleuAsnIleT 849
503 GTGGTATCTGG.....GATGCTACTCTGAGTTAAGAGTAAAAAGAGA 546
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
849 hrGlyThrlPheThrlasnAsnIlyThrlAlaAsnIleAsnIleLysgl 865
547 TATAGTGAGACCTATGGAAGTACACT.....ATAATATACAC 584
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
866 ValValIlysnleuGlnIlyAspIleAsnAsnLysglYleuAsnIleTh 882
585 TATTAAATTAAT..... 597
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
882 rThrlasnAlaSerGlyThrlGlnIlyThrlleIleAsnIlyAsnIleThra 899
598 ..GATAAGGAAATATTCAGATATGTTTACCTCAGTCAAAAGTGACGT 645
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
899 snIlysnIlyAspIleuAsnIle.....LysAsnIleLysAlaAspAla 913
646 CCGGTGAGATTT.....AACTGCGTCCCAACTGGTGGGCGACATATAT 689
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
914 GluIleGlnIleGlyAsnIleSerGlnLysglYlnuLysnleuThrl 930

```

```

690 TGGAGAAATCTGTGATATGCTTTTATGATGATATAGTACTACACA 739
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
930 eSerSerAspLysValasnl.....Thrlasn 940
740 GCAGCTCTTTGGAGATTAAGATTTCAG.....GATTAACAATCTT 777
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
940 lInleThrlleLysAlaIlyValGlnIlyGlyLyrGserAspSerSerGln 956
778 AATTCGATGCGAAATTTTATCTAAGGAAATTAATGATACACCAAGA 827
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
957 AlaGluasnAlaAsnleuThrlleGln.....ThrlYsgl 968
828 AATTCATATACCTTTGTCACCTCTCTTGCGGCT.....AAAAGTT 868
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
968 uLeuLysleuAlaIlyAspIleuAsnIleSerGlyPheAsnLysAlaGln 985
869 TAACTCCAAACAAATGGAACGTCATTAATATTTGCTGACGAGCTTCT 918
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
985 lThrlAlaLysasnIlySerAspIleuThrlleGlyAsnAlaSerGly 1001
919 GAACAACAACTGGAATAGATTAACGCTGCACCATGCCAAGATCAGTG 968
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1002 AsnAlaAspAlaLysLys.....ValThrlPheAspLysValIysas 1015
969 TCCGCTGTGTGTGGCTGGA...CGTTTGAATTTGATGCAAAAGTGA 1015
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1015 pSerLysIleSerThrlAspLysIlyAsnValThrlleAsnSerGlnVal 1032
1016 AAAATCCGAGGCTGCACAATATATGAGTATATTAATGTTACT 1059
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1032 YsThrlserAsnIlySerSerAsnAlaIlyAsnAspAsnSerThrl 1046

```

seq_name: /cgn2_6/plodata/2/laa/5B_COMB.pep:US-08-737-716-13

seq_documentation_block:

```

/ Sequence 13, Application US/08737716
/ Patent No. 5955258
/ GENERAL INFORMATION:
/ APPLICANT: Glabe BUIST
/ APPLICANT: Gerard VENEMA
/ APPLICANT: Jan KOK
/ APPLICANT: Adrianus Marinus LEDEROER
/ TITLE OF INVENTION: Process for the lysis of a culture of lactic
/ TITLE OF INVENTION: acid bacteria by means of a lysin, and uses of the resultn
/ NUMBER OF SEQUENCES: 16
/ CORRESPONDENCE ADDRESS:
/ ADDRESS: Pillsbury Madison & Sutro, L.L.P.
/ STREET: 1100 New York Avenue, N.W.
/ CITY: Washington
/ STATE: D.C.
/ COUNTRY: U.S.A.
/ ZIP: 20005-3918
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Microsoft word
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/737,716
/ FILING DATE: 22-APR-1997
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: PCT/ML95/00170
/ FILING DATE: 12-MAY-1995
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: EP 94201353.3
/ FILING DATE: 12-MAY-1994
/ INFORMATION FOR SEQ ID NO: 13:
/ SEQUENCE CHARACTERISTICS: 13:
/ LENGTH: 671 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear

```


MOLECULE TYPE: protein
 ORIGINAL SOURCE: Streptococcus faecalis
 IMMEDIATE SOURCE: CLONE: Flg. 5a (S. faecalis)
 US-08-737-716-13

alignment_scores:
 Quality: 106.00 Length: 309
 Ratio: 0.716 Gaps: 17
 Percent Similarity: 47.896 Percent Identity: 23.625

alignment_block:

US-09-839-894-9 x US-08-737-716-13 ..

Align seg 1/1 to: US-08-737-716-13 from: 1 to: 671

```

121 AGAAGCAATCTTCCCAACAAATATATA...TTAATAACCATATTAC 167
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||
309 ArgYrAlaThrAspProSerTYrAsnAlaLysLeuAsnAsnValIleTh 325
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
168 AGCATACAGTGAAGTCATCTGTATGATAGAGTACTTTTATGTT 217
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
325 rAlaTYr.....AsnLeuThrGlnTYrAspThrProSerSerGlyGly 340
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
218 TGCTCTCTCAATACACTTAATGAGACATGTCACACCATGAGATCCT 267
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
340 snThrGlyGlyGlyThrValAsnProGlyThrGlyGlySerAsnAsnGln 356
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
268 AGCATGTCATCGGTC.....ACGGTGAACA.....AA 296
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
357 SerGlyThrAsnThrTYrTYrThValLysSerGlyAspThrLeuAsnLY 373
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
297 TATPACATTCATATT.....ACGAAAAAGATTTA.... 330
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
373 sIleAlaIleGlnTYrGlyValSerValAlaAsnLeuArgSerTrpAsnG 390
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
331 ..ATAAAAAGAGCTACAATTAAAGCTATATAACAATTATTGTCAA 378
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
390 LysLeuSerGlyAspLeuIlePheValGlyLysLeuIleValLysLys 406
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
379 AGTGTACTAGCCCATCCGACCTTAACCTTAACCTCATTTTACTG 428
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
407 GLYAla.....SerGlyAsnThrGlyGlySerGlyAsnGlyLysE 420
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
429 TAATAAAACGCGGCTCAGGTGCACTTATATATTATATTCCTGCTG 478
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
420 rAsnAsnAsn...GlnSerGlyThrAsnThrTYrTYrThValLysSerG 436
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
479 GCGAAGCTAAAAATTGCTTTGGTGTGATCTGGGATGCTACTCTGAAG 528
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
436 LysAspThrLeuAsn..... 440
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
529 TTAAGACTAAAAAGACGATATAGTACCTATAGGACCTTACACTATAA 578
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
441 .....LysIleAlaIleGlnTYrGly..... 447
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
579 TATCACTATTAAATTAACTGATAAGGAAATATTACATATGTTACTCTC 628
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
448 ValThrVal.....AlaAsnLeuArgSerTrp..... 456
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
629 AGTTCAAAATGACGCTCGCTGATCTTAACTTG..... 663
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
457 .....AsnGlyIleSerGlyAspLeuIlePheValGlyLysLeu 470
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
664 .....CGTCAACTGTTGGGCGACATATATTGAGAAATATCTGT 704
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
471 IleValLysLysGlyThrSerGlyAsnThr..... 480
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
705 TGATATGTGCTTTTATGATGATATAGTACTTAACAGCAGCTCTTTGAGA 754
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
481 .....GlyGlySerSerAsnGlyLysSer..... 488
  
```

```

755 TAAGATTTCAGATACATCTTAATCTGATGGAAATTTTATCTAAG 804
    :|||:|||||:|||||:|||||:|||||:|||||:|||||
489 .....AsnAsnAsnGlnSerGlyThrAsnThrTYrTYrThIle 501
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
805 AAATAATATGATGACACCAAGAAATTTGCA.....TATCTTTGTCA.. 846
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
502 LysSerGlyAspThrLeuAsnLysIleAlaIleGlnTYrGlyValSerVa 518
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
847 .....CTTCTCTGGCGG 859
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
518 lAlaAsnLeuArgSerTrpAsnGlyIleSerGlyAspLeuIlePheAlaG 535
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
860 GTAAGTTTAATCACTCAACAAATGAGACGTCATTAATATTGCTGACGCA 909
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
535 LysGlnLysIleIleValLysGlyThrSerGlyAsnThrGlyLysSer 551
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
910 GCTTCTCTGGAACCAACTGGAATAGA 936
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
552 SerAsnGlyLysSerAsnAsnAsnGln 560
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

seq_name: /cgn2_6/plodata/2/1aa/5B.COMB.pep:US-08-728-470-9

seq_documentation_block:
: Sequence 9, Application US/08728470
: Patent No. 5928651
: GENERAL INFORMATION:
: APPLICANT: Barenkamp, Stephen J
: TITLE OF INVENTION: High Molecular Weight Surface Proteins
: TITLE OR INVENTION: of No. 5928651-Typeable Haemophilus
: NUMBER OF SEQUENCES: 10
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Shoemaker and Maltare, Ltd.
: STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
: STREET: Bldg. 1
: CITY: Arlington
: STATE: Virginia
: COUNTRY: U.S.A.
: ZIP: 22202-0286
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/728,470
: FILING DATE:
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/302,832
: FILING DATE: 16-MAR-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/US93/02166
: FILING DATE: 16-MAR-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: GB 9205704.1
: FILING DATE: 16-MAR-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: BerksStresser, Jerry W
: REGISTRATION NUMBER: 22,651
: REFERENCE/DOCKET NUMBER: 1038-633
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703) 415-0810
: TELEFAX: (703) 415-0813
: INFORMATION FOR SEQ ID NO: 9:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1338 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: US-08-728-470-9
  
```

alignment_scores:
 Quality: 101.00 Length: 418
 Ratio: 0.529 Gaps: 18
 Percent Similarity: 45.694 Percent Identity: 21.531

alignment_block:

US-09-839-894-9 x US-08-728-470-9 ..

Align seg 1/1 to: US-08-728-470-9 From: 1 to: 1338

```

61 GATGCGCATTAATTCGCCGAGATGAAGCATACATATATTTTGG 110
   |||:|||||:
817 ValThrPheAspLysVal..... 822
111 CCCGCGTACAGAGCAATCTCCCAACATATATA..TTAAATA 157
   |||:||||:|||||:|||||:
823 ...LysAspSerLysIleSerThrAspGlyHisAsnValThrLeuAsn 838
158 ACCATATT...ACAGCATACAGTGAAGTCATCTGTATGATGAGATG 204
   |||:||||:|||||:|||||:
838 erGluValLysThrSerAsnGlySerSerAsnIleGlyAsnAspAsnSer 854
205 ACTTTTATGTTGCT..... 222
   ||| ||| :|||
855 ThrGlyLeuThrIleSerAlaLysAspValThrValAsnAsnValThr 871
223 .TCTCAATATACACTTAAT.....GGAGCATGTCACCA 256
   |||:||||:|||||:|||||:
871 rSerHisLysThrIleAsnIleSerAlaAlaIleGlyAsnValThrThr 888
257 GTGAGAAATCTCAGTCATGCGTCAGCGGTAACAATATATACATTA 306
   |||:||||:|||||:|||||:
888 yseGluGlyThrThrIleAsnAlaThrThrGlySerValGluValThrAla 904
307 CAA.....TTTAC 314
   ||| |||
905 GluAsnGlyThrIleLysGlyAsnIleThrSerGluAsnValThrValThr 921
315 GCAAAAAGAGTTTATATAAAGAGCTACAAATT.....A 352
   | :|||:||||:
921 rAlaThrGluAsnLeuValThrThrGluAsnAlaValIleAsnAlaThrS 938
353 AAGGCTATTAACAATTATGTTCAAAAGTGTAACTGCCCGCCGCGCTA 402
   |||:||||:|||||:|||||:
938 erGlyThrValAsnIleSerThrLysThrGlyAspIleLysGlyGlyLe 954
403 ACACTTAACCTACGTCATTTTAACGTATAAACCCTCAGCTGC 452
   |||:||||:|||||:|||||:
955 GluSerThrSerGlyAsnValAsnIle.....ThrAlaSerGlyAs 968
453 AAGTTTA..... 459
   ||| |||
968 nThrLeuLysValSerAsnIleThrGlyGluAspValThrValThrAla 985
460 .....TATTTATATATTCCTGCTGGCAACTAAATAATTTGCCCTTT 501
   ||| ||| |||:||||:
985 spAlaGlyAlaLeuThrThrThrAlaGlySerThrIleSerAlaThrThr 1001
502 GGTGATATCTGGGATCTACTCTGAAGTTAAGCTAAAAAGACGATATAG 551
   |||:||||:|||||:|||||:
1002 GlyAsnAlaAsnIleThrThrLysThrGlyAspIleAsnGlyLysVal 1018
552 TGGACCTATGGAAGCTTACACTATAATATCTATTAATTAACGATTA 601
   |||:||||:|||||:|||||:
1018 uerSerSerGlySerValThrLeuValAlaThrGlyAlaThrLeuAla 1035
602 AGGGAATATTCAGATATGTTACTCAAGTCAAAAGTCAGCTCCGCTC 651
   |||:||||:|||||:|||||:
1035 alGlyAsnIleSerGlyAsnThrValThrThrIleThrAlaAspSerGly 1050
652 GATCTTAATCTGCGTCAACCTGGTGGGCGACATATATGGAAGAAATTC 701
   |||:||||:|||||:|||||:
1051 .....LysLeuThrSerThrValGlySerThrIleAsnGlyThrAsnS 1065

```

```

702 TGTTCATATGCTGCTTTTATGATGATATAGTACTAACAGACGCTTTGG 751
   ||| ||| |||:||||:|||||:
1065 rValThrThr.....SerSerGluSerGlyAspIleG 1076
752 AGATTAAGATTTCAGATTAACAAT.....CTTAATCTGATGGGAA 792
   |||:||||:|||||:|||||:
1076 LuGlyThrIleSerGlyAsnThrValAsnValThrAlaSerThrGlyAsp 1092
793 TTTTATCTA...AGGAAATTAATGATGACACCAAGAAATTCATATAC 839
   |||:||||:|||||:|||||:
1093 LeuThrIleGlyAsnSerAlaLysValGluAlaLysAsnGlyAlaAlaThr 1109
840 TTTTGCA.....CTTCTTGGCGGGTAAAGTTTAA 871
   |||:||||:|||||:|||||:
1109 rLeuThrAlaGluSerGlyLysLeuThrThrGluThrGlySerSerIleT 1126
872 CTCCAACAAAATGAGACGTCATTAATATGCTGACGACGCTTCTGGA 921
   |||:||||:|||||:|||||:
1126 hrSerSerAsnGlyGluThrThrLeuThrAlaLysAspSerSerIleAla 1142
922 ACAAACTGCAAT.....AGAAATTCAGCTGTCCACATGCGCAAAATCAG 965
   ||| ||| |||:||||:
1143 GlyAsnIleAsnAlaAlaAsnValThrLeuAsnThrThrGlyThrLeuThr 1159
966 TGT.....CCGCTGTTGTTGGCTCGGAGCTTTGCAATTC 1002
   |||:||||:|||||:|||||:
1159 rThrThrGlyAspSerLysIleAsnAlaThrSerGlyThrLeuThrIleAla 1176
1003 .....GATGCAAAAGTGAATAATCCGAGCTGCA..... 1032
   |||:||||:|||||:|||||:
1176 snAlaLysAspAlaLysLeuAspGlyAlaAlaSerGlyAspArgThrVal 1192
1033 .....CAATATATGCTATATTAATGTTACTTTTACACACAG 1070
   |||:||||:|||||:|||||:
1193 ValAsnAlaThrAsnAlaSerGlySerGlyAsnValThrAlaLysThrSe 1209
1071 TACT 1074
   ||| |||
1209 rSer 1210

```

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-719-641-9

seq_documentation_block:

Sequence 9, Application US/08719641

Patent No. 6218141

GENERAL INFORMATION:

APPLICANT: Barenkamp, Stephen J

TITLE OF INVENTION: High Molecular Weight Surface Proteins

TITLE OF INVENTION: of No. 6218141-Typeable Haemophilus

NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:

ADDRESSER: Shoemaker and Maltre, Ltd.

STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza

CITY: Arlington

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22202-0286

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/719,641

FILING DATE:

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/302,832

FILING DATE: 16-Sep-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US PCT/US93/02166

FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-625
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1338 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-719-641-9

alignment_scores:
Quality: 101.00 Length: 418
Ratio: 0.529 Gaps: 18
Percent Similarity: 45.694 Percent Identity: 21.531

alignment_block:
US-09-839-894-9 x US-08-719-641-9 ..
Align seg 1/1 to: US-08-719-641-9 from: 1 to: 1338

```
61 GPATCGGACATAAATTCGCGGAGATGAAGCATACATATTTTGG 110
|||||: |||||:
817 ValThrheaspIysVal..... 822
111 CCGCGGTGACAGAACGATCTTCCCAACATATATA...TTAATA 157
|||||: |||||: |||||: |||||:
823 ...LysAspSerLysIleSerThrAspGlyHisAsnValThrLeuAsn 838
158 ACCATATT...ACAGCATACAGTGAAGCATCTGTATGATAGATG 204
|||||: |||||: |||||: |||||:
838 erGIuValLysThrSerAsnGlySerSerAsnAlaGlyAsnAspAsnSer 854
205 ACTTTTATGTGTGCT..... 222
|||||: |||||:
855 ThrGlyLeuThrIleSerAlaLysAspValThrValAsnAsnAsnValThr 871
223 ...TTCACAAATACACTTAAT.....GGAGCATGTCCACCA 256
|||||: |||||: |||||: |||||:
871 rSerHisLysThrIleAsnIleSerAlaAlaAlaGlyAsnValThrThrL 888
257 GTGGAATCCTAGCAGTTCATCGGTGAGCGGTGAACAAATATACATTA 306
|||||: |||||: |||||: |||||:
888 ysgIuGlyThrThrIleAsnAlaThrThrGlySerValGIuValThrAla 904
307 CAA.....TTTAC 314
|||||: |||||:
905 GluAsnGlyThrIleLysGlyAsnIleThrSerGIuAsnValThrValThr 921
315 GGAAGAAAGAGTTTAAATAAGAGACATACAAT.....A 352
|||||: |||||: |||||: |||||:
921 rAlaThrGluAsnLeuValThrThrGluAsnAlaValIleAsnAlaThrS 938
353 AAGGCTATTAACATTAATGTTCAAAGGTGTAATGCCATCCGCGCTA 402
|||||: |||||: |||||: |||||:
938 erGIuThrValAsnIleSerThrLysThrGlyAspIleLysGlyGlyLe 954
403 ACACTTAACGAGCTCATTTTAACTGTAATATAAAGCCGCTTCACGTGC 452
|||||: |||||: |||||: |||||:
955 GluSerThrSerGlyAsnValAsnIle.....ThrAlaSerGlyAs 968
453 AAGTTTA..... 459
|||||: |||||:
968 nThrLeuLysValSerAsnIleThrGlyGluAspValThrValThrAla 985
```

```
460 .....TATTATATTCCTGCGGACATAAAAAATTTCCCTTT 501
|||||: |||||: |||||: |||||:
985 spAlaGlyAlaLeuThrThrThrAlaGlySerThrIleSerAlaThrThr 1001
502 GGTGTATCTGGATGCTACTCTGACGTTAAGATTAAGAAAGACATATAG 551
|||||: |||||: |||||: |||||:
1002 GlyAsnAlaAsnIleThrThrLysThrGlyAspIleAsnGlyLysValGI 1018
552 TGAGACCTATGGAACCTTACATTAATATCAGTATTAATTAATCACTGATA 601
|||||: |||||: |||||: |||||:
1018 userSerSerGlySerValThrLeuValAlaThrGlyAlaThrLeuAlav 1035
602 AGGCAATATTTCAGATATGTTACCTCAGTTCAAAAGTACGCTCGCTC 651
|||||: |||||: |||||: |||||:
1035 alGIyAsnIleSerGlyAsnThrValThrIleThrAlaAspSerGly... 1050
652 GATCTTAACCTTGCGTCCAGCTGTGGGCGACATATATTGAGAAATTC 701
|||||: |||||: |||||: |||||:
1051 .....LysLeuThrSerThrValGIySerThrIleAsnGlyThrAsnSe 1065
702 TGTGATATGTCTTTATGATGATATAGTACTACACAGCATCTTTGG 751
|||||: |||||: |||||: |||||:
1065 rValThrThr.....SerSerGlnSerGlyAspIleG 1076
752 AGATAAGATTTCAGCATACAT.....CCTAAATCTGATGCGCAA 792
|||||: |||||: |||||: |||||:
1076 IuGIyThrIleSerGlyAsnThrValAsnValThrAlaSerThrGlyAsp 1092
793 TTTATCTA...AGAAATTAATGATGACACCAAGAAATTCGATATAC 839
|||||: |||||: |||||: |||||:
1093 LeuThrIleGlyAsnSerAlaLysValGIuAlaLysAsnGlyAlaAlaThr 1109
840 TTTGTCA.....CTTCTCTGCGGTAAGTTTAA 871
|||||: |||||: |||||: |||||:
1109 rLeuThrAlaGluSerGlyLysLeuThrThrGlnThrGlySerSerIleT 1126
872 CTCACAAATATGACGTCATTAATATGCTGACGCCAGCTTCTGTGAA 921
|||||: |||||: |||||: |||||:
1126 hrSerSerAsnGlyGlnThrThrLeuThrAlaLysAspSerSerIleAla 1142
922 ACAACGTGGAAT.....AGATATACAGCTGTACCATGTCAGAAATCAG 965
|||||: |||||: |||||: |||||:
1143 GlyAsnIleAsnAlaAlaAsnValThrLeuAsnThrThrGlyThrLeuThr 1159
966 TGTT.....CCGCTGTGTGTGCTGCGACGTTTGCATTTG 1002
|||||: |||||: |||||: |||||:
1159 rThrThrGlyAspSerLysIleAsnAlaThrSerGlyThrLeuThrIleA 1176
1003 .....GATGCAAAAGTGAAATCCCGAGCTGCA..... 1032
|||||: |||||: |||||: |||||:
1176 snAlaLysAspAlaLysLeuAspGIyAlaAlaSerGlyAspArgThrVal 1192
1033 .....CAATATATGGTAAATATTAATTTACTTTCACACCAAG 1070
|||||: |||||: |||||: |||||:
1193 ValAsnAlaThrAsnAlaSerGlySerGlyAsnValThrAlaLysThrSe 1209
1071 TAGT 1074
|||||: |||||:
1209 rSer 1210
```

seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-617-697-9

seq_documentation_block:
Sequence 9, Application us/08617697
Patent No. 5977336
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteins
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mattare, Ltd.

STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/617,697
FILING DATE: 01-Apr-1996
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 05-Oct-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-557
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1599 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-617-697-9
alignment_scores:
Quality: 101.00 Length: 418
Ratio: 0.529 Gaps: 18
Percent Similarity: 45.694 Percent Identity: 21.531
alignment_block:
US-09-839-894-9 x US-08-617-697-9 ..
Align seg 1/1 to: US-08-617-697-9 from: 1 to: 1599
61 GATCGGACGATTAATTCGCCGAGATGAAGCATTAATATTTTGG 110
|||||
1077 ValThrPheAspLysVal..... 1082
111 CCGCGGACAGAGCAATCTCCCAACATTAATA..TTAATA 157
|||||
1083 ..LysAspSerLysIleSerThrAspGlyHisAsnValThrLeuAsn 1098
158 ACCAATAT...ACAGCATACAGTGAAGTCATCTGTATGATAGATG 204
|||||
1098 erGluValLysThrSerAsnGlySerSerAsnLagLysAsnAspAsnSer 1114
205 ACCTTTTATGTTGCT..... 222
|||||
1115 ThrGlyLeuThrIleSerAlaLysAspValThrValAsnAsnValThr 1131
223 ..TCGCAATACACTTAAT.....GGAGCATGTCAACCA 256
|||||
1131 rSerHisLysThrIleAsnIleSerAlaAlaLagLysAsnValThrThr 1148
257 GTGAGATCCTACAGTCATCGTCAGCGGTGAACAATATATACATTA 306
|||||
1148 yseGluGlyThrThrIleAsnAlaThrThrGlySerValGluValThrAla 1164
307 CAA.....TTTAC 314
||

1165 GluAsnGlyThrIleLysGlyAsnIleThrSerGlnAsnValThrValTh 1181
315 GGAAGAAAGAGTTTATATAAAGAGCGTACAAAT.....A 352
|||||
1181 rAlaThrGluAsnLeuValThrThrGluAsnAlaValIleAsnAlaThrS 1198
353 AAGCGTATAAACAATATGTTCCAAAGTTTAACGCCCATCCGCCCTA 402
|||||
1198 erGlyThrValAsnIleSerThrLysThrGlyAspIleLysGlyGlyIle 1214
403 ACACCTTAACACCTCATTTTAACGTAAATAAAGCGGCTTCAGATGC 452
|||||
1215 GluSerThrSerGlyAsnValAsnIle.....ThrAlaSerGlyAs 1228
453 AAGTTTA..... 459
|||||
1228 nThrLeuLysValSerAsnIleThrGlyGlnAspValThrValThrAla 1245
460TATTTATATATTCCTGCGCGCACTAAATAATTTGCCCTTT 501
|||||
1245 spAlaGlyAlaLeuThrThrThrAlaGlySerThrIleSerAlaThrThr 1261
502 GGTGGTATCGGATCGTACTCTGAAGTTAAGCTAAAAAGAGATATAG 551
|||||
1262 GlyAsnAlaAsnIleThrThrLysThrGlyAspIleAsnGlyLysValG 1278
552 TGAACCTATGAACTTACCTCACTAATATACACTTAATTAATTAACGAT 601
|||||
1278 uSerSerSerGlySerValThrLeuValAlaThrGlyAlaThrLeuAla 1295
602 AGGAAATATTCAGATATGTTACTCCTCAGTTCAAAAGTGACGCTCCGC 651
|||||
1295 alGlyAsnIleSerGlyAsnThrValThrIleThrAlaAspSerGly... 1310
652 GATCTTAACCTGCGTCCCACTGGTGGGCGACATATTTGGAAGAAATTC 701
|||||
1311LysLeuThrSerThrValGlySerThrIleAsnGlyThrAsnSe 1325
702 TGTGATATGTCGCTTTATGATGATATAGTACTACAGACAGCGCTTTGG 751
|||||
1325 rValThrThr.....SerSerGlnSerGlyAspIleG 1336
752 AGATAAGATTTCAGATTAACAAT.....CCTAATCTGATGGGAAA 792
|||||
1336 luGlyThrIleSerGlyAsnThrValAsnValThrAlaSerThrGlyAsp 1352
793 TTTTATCTA...AGAAATTAATGATGACACCAAGAAATTCATATAC 839
|||||
1353 LeuThrIleGlyAsnSerAlaLysValGluAlaLysAsnGlyAlaAlaTh 1369
840 TTTGTCA.....CTTCTGTGGCGGTAAAGTTTAA 871
|||||
1369 rLeuThrAlaGluSerGlyLysLeuThrThrGlnThrGlnThrSerIleT 1386
872 CTCCAACAAATGAGACGTCATTAATATGCTGACGACGCTTCGTGAA 921
|||||
1386 hrSerSerAsnGlyGlnThrThrLeuThrAlaLysAspSerSerIleAla 1402
922 ACAAACTCGAAT...AGAAATTACAGCTGTACCATGCGCAAAATCAG 965
|||||
1403 GlyAsnIleAsnAlaAlaAsnValThrLeuAsnThrThrGlyThrLeu 1419
966 TGT.....CCGGTGTGTGGTGGCTGAGCGTTTCATTTG 1002
|||||
1419 rThrThrGlyAspSerLysIleAsnAlaThrSerGlyThrLeuThrIleA 1436
1003GATGCAAAAGTGAAGAAATCCGAGGCTGGA..... 1032
|||||
1436 snAlaLysAspAlaLysLeuAspGlyAlaAlaLaserGlyAspArgThrVal 1452
1033CAATATATGGTAAATATATATGTTACTTTCCACACCAAG 1070
|||||
1453 ValAsnAlaThrAsnAlaSerGlySerGlyAsnValThrAlaLysThrSe 1469


```

CORRESPONDENCE ADDRESS:
ADDRESS: Flehr, Hohbach, Test, Albritton & Herbert
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10661A
FILING DATE: 16-AUG-1995
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/236,791
FILING DATE: 25-AUG-1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Treacartin, Richard F.
REGISTRATION NUMBER: 31,801
REFERENCE/DOCKET NUMBER: FP-59941/RFT
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEO ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1848 amino acids
TYPE: amino acid
TOPOLOGY: unknown
PCT-US95-10661A-6

alignment_scores:
Quality: 99.00 Length: 325
Ratio: 0.593 Gaps: 20
Percent Similarity: 51.385 Percent Identity: 23.385

alignment_block:
US-09-839-894-9 x PCT-US95-10661A-6 ..
Align seg 1/1 to: PCT-US95-10661A-6 from: 1 to: 1848

55 TTTGGCTGATCGGAGATAAATTCGCCGAGATGAACATTAATATAT 104
||| : : : : : : : : : : : : : : : : : : : : : : : :
753 PheLysAlaThrThrMetAsnValThrGlyAsnAlaSerLeuTyrSer.. 768
105 TTTTGGCCCGCGTGACGAGAACGANTCTCCCAACATAATATATTAA 154
||| : : : : : : : : : : : : : : : : : : : : : : : :
769 .. : : : : : : : : : : : : : : : : : : : : : : : :
155 ATTAACCATATTACAGCATACAGTGAAGT.....CATACTCTGTATGAT 198
: : : : : : : : : : : : : : : : : : : : : : : : : :
777 eirAsn...IleThrAlaSerAsnAsnAlaGlnValHisIleGlyTyrLys 792
199 AGGATGACTTTTATGTTGTTCTCT..... : : : : : : : : : : 225
: : : : : : : : : : : : : : : : : : : : : : : : : :
793 ThrGlyAspThrValCysValArgSerAspTyrThrGlyTyrValThrCys 809
226 ..CAACAATACA...CTTAATGAGCATGTCCACACGAGTAATCTTACGA 271
||| : : : : : : : : : : : : : : : : : : : : : : : :
809 sHisAsnSerAsnLeuSerGluLysAlaLeuAsnSerPheAsnPro.... 824
272 GTTCAATCGGTGACGGGTGAACAATATAACA.....TTACAAATTAGC 315
: : : : : : : : : : : : : : : : : : : : : : : : : :
825 ..ThrsAsnLeuArgGlyAsnValAsnLeuThrGluAsnAlaSerPheThr 840
316 GAAAAAAGAGTTTAATAAAGAGAGACTCAAAATTAAGGCTATAACA 365
: : : : : : : : : : : : : : : : : : : : : : : : : :
841 LeuGlyLysAlaAsnLeuPheGlyThrIleGlnSerIleGlyThrSerGln 857

```

```

366 ATTAATGTTCAAAAGTGTAACTGCCATCGGCTTAACATTAACACG 415
||| : : : : : : : : : : : : : : : : : : : : : : : :
857 nValAsnLeuLysGlu...AsnSerHisThrPheLeuThrGlyAsnSer.. 872
416 CTCAATTTAACTGTAAATAA...AACGGGCTTCAGGTGCAAGTTTAT 462
||| : : : : : : : : : : : : : : : : : : : : : : : :
873 .. : : : : : : : : : : : : : : : : : : : : : : : :
463 TTATATATTCCTGCTGGCAACTAAATAATTGGCTTTGGTGTATCTG 512
: : : : : : : : : : : : : : : : : : : : : : : : : :
884 IleHisLeuAsnAlaGlnAsnAspAlaAsn..... : : : : : 893
513 GGATGCTACTCTGAGATTAAAGTAAAGACGATATAGTGAGACCTATG 562
||| : : : : : : : : : : : : : : : : : : : : : : : :
894 .. : : : : : : : : : : : : : : : : : : : : : : : :
563 GAACCTTCACTATAATATACATTTAAATTAAGTGAAGGAATATTT 612
: : : : : : : : : : : : : : : : : : : : : : : : : :
899 snThrLeuThrValAsn.....SerLeuSerGlyAsnGlySerPhe 912
613 CAGATATGGTTACCTCAGTTCAAAAGTGACGCT...CGCGTCGATCTTAA 659
||| : : : : : : : : : : : : : : : : : : : : : : : :
913 TyrTyrThrPalaAspPheThrAsnAsnLysSerAsnLysValValAs 929
660 CTTCGCTCAACTGCTGGGGGCACATAT.....ATTGAA 694
||| : : : : : : : : : : : : : : : : : : : : : : : :
929 nLysSerAlaThrGlyAsnPheThrLeuGlnValAlaAspLysThrGly 946
695 GA...AATTCGTGATATGCTTTTATGATGATGATTAAGTACTAACAAC 741
: : : : : : : : : : : : : : : : : : : : : : : : : :
946 IuProAsnHisAsnGlnLeuThrLeuPheAspAlaSerAsnAlaThrArg 962
742 AGCTCTTGGAGATAAGATTTCAGATTAACAAT...CCTAAATCTGATGG 788
: : : : : : : : : : : : : : : : : : : : : : : : : :
963 AsnAsnLeuGlnValThrLeuAlaAsnGlySerValAspArgGlyAlaThr 979
789 GAATTTTATCTAAGAAATAATATGATGACACCCAAAGAAATTGCATATA 838
||| : : : : : : : : : : : : : : : : : : : : : : : :
979 pLysTyrLysLeuArgAsnValAsnGly....ArgTyrAspLeuTyrAla 994
839 CTTCGTCACTTCCTTGGCGGGTAAAGTTTAACCTCAACAATATGACG 888
: : : : : : : : : : : : : : : : : : : : : : : : : :
994 snProGluValGlnLysArgAsnGlnThrValAspThrThrAsnIleThr 1010
889 TCATTAAAT.....ATTGCTGACGACGCTTCCTGGAACAACATCGAA 932
: : : : : : : : : : : : : : : : : : : : : : : : : :
1011 ThrProAsnAspIleGlnAlaAspAlaProSerAlaGlnSerAsnAsnGln 1027
933 TAGAATTACAGCTGTGACCATGCGCA 957
: : : : : : : : : : : : : : : : : : : : : : : : : :
1027 uGlnIleAlaArgValGlnThrPro 1035

seq_name: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:US-08-409-995-4
seq_documentation_block:
; Sequence 4, Application us/08409995
; Patent No. 5646259
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen I.
; APPLICANT: St. Geme III, Joseph W.
; TITLE OF INVENTION: Haemophilus Adhesion Proteins
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

```

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/409,995
FILING DATE: 24-MAR-1995
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Silva, Robin M.
REGISTRATION NUMBER: 38,304
REFERENCE/DOCKET NUMBER: A-61053/RFT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1912 amino acids
TYPE: amino acid
STRANDEDNESS: double
TOPOLOGY: unknown
US-08-409-995-4

```

```

alignment_scores:
Quality: 94.50 Length: 310
Ratio: 0.602 Gaps: 14
Percent Similarity: 50.645 Percent Identity: 19.677

alignment_block:
US-09-839-894-9 x US-08-409-995-4 ..
Align seg 1/1 to: US-08-409-995-4 from: 1 to: 1912

```

```

232 ACACCTTAATGAGCATGTCACCAACGAGTAGAGATCCTAGCAGTTCAATCGGT 281
|||||
805 ThrilegylgYAsnThrProthr.....GlyGlyThrThrAl 817
282 CAGCGGTAAACAAATATACATTTACATTACGAGAAAAAGAACTTTAA 331
|||||
817 aThrProlySvalasnThr.....SerThrAlaspolyleuasnph 833
332 TAAAGAGAGACCTACAAATTAAGGCTTAACATTAATTTGTTCAAAAGT 381
|||||
833 lalysglThrAlaspaLaserglySerlySasnValTylleuysglY 849
382 GTTAACCTC.....CCATCCGCGCTTAACACTTAACCTGACCTCA 419
|||||
850 lIeaIaThrThrleuThrGluProserAla..GlyAlalysSerSerH 865
420 TTTTACTGTATATAAACCGG.....GCTTCAGGTGCAAGTTTAT 460
|||||
865 sValaspleuasnValaspaIaThrlySerSerasnAlaIaSerIleG 882
461 ATTATATATTCCTGCT.....GGCAACTAAATAATTTG 495
|||||
882 lAsasValleuThrGlyAlaGlyTTrpaenIleGInglyasnGlyasnVal 898
496 CTTTGTGCTATCTGGATGCTACCTGGAAGTTAAGTAAAAAGACG 545
|||||
899 AspTylValAlaThrTyrAspThrVal.....AspPh 909
546 ATATAGTGAGACCTTGAACCTTACATTAATATTCACCTATTAATTAA 595
|||||
909 eThraspasSerThrGlyThrThrThrValThrValThrGlnIlySAla 926
596 CTGATAAGGA..... 606
|||||
926 spGlylysglyAlaAspVallyIleGlyAlaIyThSerValIlelys 942
607 .....AATATTCAGATATGTTTACCTCAGTTCAAAAGTGACGCTGGGT 650
|||||
943 AspHisasnGlylyLeuPherThrGlylySaspleuIySaspaIaAsnAs 959

```

```

651 CGATCTTAACCTG.....CGTCCACCTGGTGGGCACAT 685
|||||
959 nGlyAlaThrValSerGluAspAspGlylySasphThrGlyThrGlyLeu 976
686 AATATGGAAGAATTCGTGTGATATGCTTTTATGATGATAT..... 729
|||||
976 aThrAlaIlyThrValIleAspaIaValasnLySerGlyTTrpaRyAl 992
729 ..... 729
993 ThrGlyGlyAlaThrAlaGluThrGlyAlaThrAlaValaIasnAlaG 1009
730 .....AGTACTACAGACGCTTTGGAGTAAATTTGAGATTAACA 772
|||||
1009 yAsnAlaGluThrValThrSerGlyThrSerValasnPhelySasnGly 1026
773 AT.....CCTAAATCTGATGAGGAAATTTATCTAAG 804
|||||
1026 snAlaThrThrAlaThrValSerlySaspaSnglyAsnIleasnValys 1042
805 .....AAATTAATGATGACACCAAGA 827
|||||
1043 TyrAspValasnValGlyAspGlyleuysIleGlyAspAspLylySyl 1059
828 AATGCATATACTTGTCACTT...CTCTGGCGGGTAAATTTTAACTC 874
|||||
1059 eValAlaAspThrThrThrThrThrValThrGlyGlyValSerValP 1076
875 CAAACAATGAGACGTCATTAAT.....ATTGCTGACCGCAGCT 912
|||||
1076 roAlaGlyAlaIasnSerValaIasnAsnIlySlyleuValaIasnAlaGlu 1092
913 TCTCTGAAACAACTGGAATGAAATTAACA 942
|||||
1093 GlyleuAlaThrAlaIeuaIasnIleuSer 1102

```

seq_name: /cgn2_6/ptodata/2/1aa/6A_COMB.pep:US-08-685-467-4

```

seq_documentation_block:
Sequence 4, Application US/08685467
Patent No. 6060059
GENERAL INFORMATION:
APPLICANT: St. Gene III, Joseph W.
TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESSES:
ADDRESSER: Flehr, Hohbach, Test, Albritton & Herbert
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/685,467
FILING DATE: 22-JUL-1996
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/409,995
FILING DATE: 24-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Silva, Robin M.
REGISTRATION NUMBER: 38,304
REFERENCE/DOCKET NUMBER: A-61053-2/RFT/RMS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299

```


This Page Blank (uspto)

This Page Blank (uspto)

349 TyrSerGly1LeuAsnValThrPheThrProSerSerSerLeu 364
 seq_name: pir2:AE0541

seq_documentation_block:

probable fimbrial protein tcfD [imported] - Salmonella enterica subsp. enterica serovar
 C:Species: Salmonella enterica subsp. enterica serovar Typhi
 A:Note: This species has also been called Salmonella typhi
 C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 09-Nov-2001
 C:Accession: AE0541
 R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
 K.; Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
 S.; Moule, S.; O'Gaora, P.
 Nature 413, 848-852, 2001
 A:Authors: Park, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
 A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
 A:Reference number: AB0502; PMID:11677608
 A:Accession: AE0541
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-359 <PAR>
 A:Cross-references: GB:AL513382; PIDN:CAD08773.1; PID:916501589; GSPDB:GN00176
 C:Genetics:
 A:Gene: tcfD

alignment_scores:

Quality: 275.50 Length: 303
 Ratio: 1.514 Gaps: 11
 Percent Similarity: 60.066 Percent Identity: 29.703

alignment_block:
 US-09-839-894-9 x AE0541 ..

Align seg 1/1 to: AE0541 from: 1 to: 359

```

208 TTTTATGTTGTTCTTCACATACCTTAATGAGCATGTCACACG 257
    ::::||||| ::::||||| ::::||||| ::::|||||
79 TrpValcysArgSerAsnArgAsnGluAsnGluGlyAlaCysGluGlu 95
    ::::||||| ::::||||| ::::||||| ::::|||||
258 TCGAATCTCAGCATTCGCGTCAGCGGTAACCAATATATACATTAC 307
    ::::||||| ::::||||| ::::||||| ::::|||||
95 rHsLeuValAlaTrpTrpTyrAlaPheGlyAlaTyrSerIleArgLeu 112
    ::::||||| ::::||||| ::::||||| ::::|||||
308 AATTACGGAAGAAAGATTATATATAAGAGACTCAACAAATAAGGC 357
    ::::||||| ::::||||| ::::||||| ::::|||||
112 rPheArgGluGlnIleSerHisAlaGlnIleThrLeu..... 124
    ::::||||| ::::||||| ::::||||| ::::|||||
358 TATTAACATATATGTTCAAAAGCTTAAC.....TGCCATCGGCGCT 401
    ::::||||| ::::||||| ::::||||| ::::|||||
125 .....IleLeuLeuGlySerValArgAspAlaCysTyrThrGly 138
    ::::||||| ::::||||| ::::||||| ::::|||||
402 AACACTTAACAGCTCATTTTAACTGTAATAAAACGGCGCTCA.... 447
    ::::||||| ::::||||| ::::||||| ::::|||||
138 I.....IleAsnMetAlaAlaAlaAlaAlaCysG 147
    ::::||||| ::::||||| ::::||||| ::::|||||
448 .....GTCGCAAGTTATATATATATATATTCCTGCGCACTAAATAA 492
    ::::||||| ::::||||| ::::||||| ::::|||||
147 IntPglYArgSerLeuAlaArgIleProSerGluGluLeuAlaLys 163
    ::::||||| ::::||||| ::::||||| ::::|||||
493 TTGCCTTTGGTGTATCTGGAGTCTGTAAGTTA...AGACTAAA 539
    ::::||||| ::::||||| ::::||||| ::::|||||
164 IleProThrSerGlyThrTrpLysAlaThrLeuValLeuAspTyrLeu 180
    ::::||||| ::::||||| ::::||||| ::::|||||
540 AAGACATATAGTAGAGACTATGGAACCTTACACTATATAATCATATTA 589
    ::::||||| ::::||||| ::::||||| ::::|||||
180 nTrpGlyGlyAspAspProLeuGlyThrSerThrTrpAspIleThrLeu 197
    ::::||||| ::::||||| ::::||||| ::::|||||
590 AATTACAGAT.....AAGGAAATATTCAGATATGTTACTCTGCTC 633
    ::::||||| ::::||||| ::::||||| ::::|||||
197 snValThrSphHisPheAlaGluAsnAlaAlaIleTyrPheProGlnPhe 213
    ::::||||| ::::||||| ::::||||| ::::|||||

```

```

634 ...AAAGTACGCGTCGTCGATCTTAACCTTGCCTCAACCTGTGGGG 680
    ::::||||| ::::||||| ::::||||| ::::|||||
214 GlyThrAlaThrProArgValAspLeuAsnLeuHisArgMetAsnIle 230
    ::::||||| ::::||||| ::::||||| ::::|||||
681 CACATATATGGAAGAAATCTGTTGATATGTCCTTTATGATGATATA 730
    ::::||||| ::::||||| ::::||||| ::::|||||
230 rGlnMetSerGlyArgAlaAsnLeuAspMetCysLeuTyrAspGly... 246
    ::::||||| ::::||||| ::::||||| ::::|||||
731 GTACTAACGACGACCTTTGGAGATAGATTAAGATTACAGATACCTTAA 780
    ::::||||| ::::||||| ::::||||| ::::|||||
246 lValAlaLysAlaArgSerLeuGlnMetLysIleGluLysSerAsnLys 262
    ::::||||| ::::||||| ::::||||| ::::|||||
781 TCGATGAGGAAATTTATCTAAGAAATTAATGATACACCAAGAAAT 830
    ::::||||| ::::||||| ::::||||| ::::|||||
263 GlyThrGlyPheGlnValIleLysSerAspSerAlaAspThr..... 277
    ::::||||| ::::||||| ::::||||| ::::|||||
831 TGCATATACTTTGTCACCTTCTCTGGCGGTAAGATTAACCTCCACAA 880
    ::::||||| ::::||||| ::::||||| ::::|||||
277 eAspTyrAlaValSerMetAsnTyrGlyArgSerIleProValThrA 294
    ::::||||| ::::||||| ::::||||| ::::|||||
881 AATGACGTCATTAATATTTGTCAGACGACCTTCTGGAACCAACTG 930
    ::::||||| ::::||||| ::::||||| ::::|||||
294 rGluValAlaGlnPheSerLeuAspAsnValAspLysAlaAlaThrArg... 309
    ::::||||| ::::||||| ::::||||| ::::|||||
931 AATGAAATTACAGCTGTACACCGCAAGAAATCAGTTCGCGTGTGTG 980
    ::::||||| ::::||||| ::::||||| ::::|||||
310 .....ProValValLeuProGlyGlnArgGlnAlaValArgCys 322
    ::::||||| ::::||||| ::::||||| ::::|||||
981 TTGGCGCTGACGCTTTCATATGATGACAAA.....GTGGAATATC 1021
    ::::||||| ::::||||| ::::||||| ::::|||||
322 sValProValProLeuThrLeuThrGlnProPheAsnIleArgGlu 339
    ::::||||| ::::||||| ::::||||| ::::|||||
1022 CCGAGCGTGGACAAATATATGCGTAATATTAAGTTACTTTCACACCA 1071
    ::::||||| ::::||||| ::::||||| ::::|||||
339 ysArgSerGlyGlyArgGlnGlyThrLeuThrValThrMetLeuMetGly 355
    ::::||||| ::::||||| ::::||||| ::::|||||
1072 AGTCAACA 1080
    ::::||||| ::::||||| ::::||||| ::::|||||
356 ThrGlnThr 358
    ::::||||| ::::||||| ::::||||| ::::|||||

```

seq_name: pir2:G90975

seq_documentation_block:

probable factor [imported] - Escherichia coli (strain 0157:H7, substrain RIMD 0509952
 C:Species: Escherichia coli
 C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
 C:Accession: G90975
 R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.
 gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
 DNA Res. 8, 11-22, 2001
 A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli 0157:H7
 A:Reference number: A99629; MUID:21156231; PMID:11258796
 A:Accession: G90975
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-1335 <NAV>
 A:Cross-references: GB:BA000007; PIDN:BA836198.1; PID:913362243; GSPDB:GN00154
 C:Experimental source: strain 0157:H7, substrain RIMD 0509952
 C:Genetics:
 A:Gene: Ecs2775

alignment_scores:

Quality: 120.00 Length: 392
 Ratio: 0.628 Gaps: 15
 Percent Similarity: 48.724 Percent Identity: 20.408

alignment_block:
 US-09-839-894-9 x G90975 ..

Align seg 1/1 to: G90975 from: 1 to: 1335

37 TCTTCAGTACTTTTACATTGCTGTATCGGACGATAAATCCGAGACA 86

```

||||:  :: |||:|||||:  ::  ::  ::
746 SerAlaLysIleAlaThrLeuSerAlaSerAsnGlyValLeuAlaAs 762
87  TGAAGCATTAATAATTTTGGCCCGCGTGACGAGAAATCTCC 136
   |||:  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
762 nGluAsnAlaAlaSerThrValSerValAsnValAlaAspGluLys 778
137 CCAACATATATATTAATACCATTTACAGCATAC 177
   |||:  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
779  ....AsnProIleAsnAspHisThrValThrPheAlaValLeuSer 792
178 GAAATCATCTGTATGATGATGATGATGATGATGATGATGATGAT 227
   |||:  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
793 GlySerAlaThrSerPheAsnAsnGlnAsn 805
228 CAATACCTTAATGAGATGTCACACCATGAGAAATCTAGCAT 277
   |||:  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
805 sThrAspValAsnGlyLeuAla...ThrPheAspLeuLysSerLysG 821
278 CGGTACAGCGGTGAACAAATATATACATTTACGAAATAAGAGT 327
   |||:  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
821 nGluAspAsnThrValGluValThrLeuGluAsnGlyValLysGlnThr 837
328 TTAATA 333
   |||:  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
838 LeuIleValSerPheValGlyAspSerSerThrAlaGluValAspLeuG 854
334  ....AAAGAGAGCTACAATTAAGCTAT...AAACATTAATGCT 373
   |||:  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
854 nLysSerLysAsnGluValAlaAlaAspGlyAsnAspSerAlaThrMet 871
374 TCAAACTGTAACCTGCCATCGCGCTTAACATTAACATGAGCTATTT 423
   |||:  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
871 hTrAlaThrValArgAspAlaLysGlyAsnLeuLeuAsnAspValLysVal 887
424 AACTGTAATAAAACGGCGCTGAGTGAAGTTATATATATATATCC 473
   |||:  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
888 ThrPheAsnValAsnSerAlaAlaAlaLysLeuSerGlnThrGluValAs 904
474 TCGTGGCGAACTAAATAATTTGCTTTGCTGATCTGGAGCTCATCTC 523
   |||:  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
904 nSerHisAsp 914
524 TGAAGTTAAGTAAAGACGATATGATGAGACCTTGAACCTTACACT 573
   |||:  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
914 euvhr 923
574 ATAAATTCATAATTAATTAAGTAAAGAAATTTGATGATATGTT 623
   |||:  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
924 ValThrAlaSerValSerSerGlySerGlnAlaAsnGlnValIlePhe 940
624 ACCTCACTCAAAAGTGCAGCTGCGCTGATGATGATGATGATGATG 673
   |||:  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
940 eIleGlyAspGlnSerThrAlaIleuThrLeuSerVal...ProSerG 956
674 GTGGGGCCACATATATGGAAGAAATCTGTGATATGCTCTTATGAT 723
   |||:  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
956 LysPheIleThrVal 960
724 GGATATGATCTTAACAGACGCTTTGGAG 960
   |||:  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
961  ....ThrAsnThrAlaProLeuHisMetThrAlaThrLeuGlnAs 974
762 TCAGATTAACAATCTTAATCTGATGGAAATTTTATCTAAGAAATTA 811
   |||:  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
974 pLysAsnGlyAsnProLeuLysAspLysGluIleThrPheSerValPro 991
812 ATGATGACACCAAGAAATGATATATCTTGTACATCTCTCTGGCGGT 861
   |||:  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
991 snAsp 1003
862 AAAGTTTAATCTCAACAAATGAGAGCTCATTA 894
   |||:  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::

```

```

1004 LysGlyMetThrAspSerAsnGlyThrAlaIleAlaSerLeuThrGly 1020
895  ....AATATG 901
1020 rLeuAlaGlyThrHisMetIleThrAlaArgLeuAlaAsnSerAsnVal 1037
902 CTGACGAGCTTCTCTGAAACAAACATGAAATGAAATTAACGCTGACC 951
   |||:  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
1037 eAspThrGlnProMetThrPheValAlaAspLysAspArgAlaValVal 1053
952 ATGCCAGAAATCAGTGTCCGGTGTGTGCTGCA 995
   |||:  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
1054 ValLeuGlnThrSerLysAlaGluIleGlyAsnGlyValAspGluThr 1070
996 GCATTTGATGCAAAAGTGGAAATCCGAGCGCTGACATATATGGCTA 1045
   |||:  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
1070 rThrLeuThrAlaThrValLysAspPro...PheAspAsnValLysA 1086
1046 ATATTAATGTTACTTTCACACCAAGT 1071
   |||:  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
1086 snLeuSerValValPheArgThrSer 1094
seq_name: p1r2:E85822
seq_documentation_block:
probable_invasin 23135 [imported] - Escherichia coli (strain O157:H7, substrain EDL93
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: E85822
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamoustis, K.; Apoda
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: AB5480; M01D:21074935; PMID:11206531
A:Accession: E85822
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2660 <SNO>
A:Cross-references: GB:AE005174; NID:q12516151; PIDN:AA657041.1; GSPDB:GN00145; UMGCP:
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: 23135
alignment_scores:
quality: 120.00 length: 392
ratio: 0.628 gaps: 15
Percent Similarity: 48.724 Percent Identity: 20.408
alignment_block:
US-09-839-894-9 x E85822 ..
Align seg 1/1 to: E85822 from: 1 to: 2660
37 TCTTCAGTACTTTTACATTTGCTATGCGAGCATTAATTCGCCGAGA 86
   |||:  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
737 SerAlaLysIleAlaThrLeuSerAlaSerAsnGlyValLeuAlaAs 753
87  TGAAGCATTAATAATTTTGGCCCGCGTGACGAGAAATCTCC 136
   |||:  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
753 nGluAsnAlaAlaSerThrValSerValAsnValAlaAspGluLys 769
137 CCAACATATATATTAATACCATTTACAGCATAC 177
   |||:  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
770  ....AsnProIleAsnAspHisThrValThrPheAlaValLeuSer 783
178 GAAATCATCTGTATGATGATGATGATGATGATGATGATGATGATG 227
   |||:  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
784 GlySerAlaThrSerPheAsnAsnGlnAsn 796
228 CAATACCTTAATGAGATGTCACACCATGAGAAATCTAGCAT 277
   |||:  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
796 sThrAspValAsnGlyLeuAla...ThrPheAspLeuLysSerLysG 812

```



```

640 GAC...CTCGCTGATCTTACTTGGT.....CCAACTGG 674
||| :||||| :|||
340 AspheValasnLeuGlnSerGlnLeuValSerLeuProThr 356
||| :||||| :|||
675 TGGGGCACAATATATGGAAGAAATCTGTTGATATGTC..... 714
||| :||||| :|||
356 rlyGlnSerGlnLeuSerSerSerSerIleTyrLysCysAspAsnSer 373
||| :||||| :|||
715 .....TTTATGATGATATAGTACTACAGCAGCTCTTGGAG 753
||| :||||| :|||
373 lalIethrAsnIleTyrSerIlePheGlyThrAsnAsnPheThrLeuPro 389
||| :||||| :|||
754 .....ATAAGATTTCAGATACAAATCC 776
||| :||||| :|||
390 SerGlnProAlaGlnIleAlaAsnMetIleGlyTyrGlyValAsnGlyThr 406
||| :||||| :|||
777 TAAATCTGATGGAAA..... 792
||| :||||| :|||
406 rAsnThr...GlyLysIleLeuThrAspTyrAlaValProThrThrPheA 422
||| :||||| :|||
793 .....TTTATCTAAGCAAAATATAATGATGACACCAAGAAATGCAATACT 840
||| :||||| :|||
422 snTyrThrIleLysAsnAsnLysAspThr.....IleSerAlaThr 436
||| :||||| :|||
841 TTGCTACTCTCTGGCGGTAAGTTTAACCTCCACAATGCAAGCTC 890
||| :||||| :|||
437 IleSerTyrAspLysAla...AsnSerLeuAsnGlnLeuAspValThrAl 452
||| :||||| :|||
891 ATTAATATATGCTGACGACGCTCTCGAAACAACACTGAAATGA...A 937
||| :||||| :|||
452 aThrThrValAlaLysSerAlaSerThrSerGlnSerSerSerAla 469
||| :||||| :|||
938 TTACAGCTGCACCATGCCAGAAATCAGTGTCCGGTGTGTGGCT 987
||| :||||| :|||
469 eufThrSerSerThrSerProSerSerSerThrGlySerSerSerThr 485
||| :||||| :|||
988 GGACGCTTTCATGATGCAAAAGTGAAATCCCGAGCTGCACATA 1037
||| :||||| :|||
486 Gly.....SerSerSerAlaSerSerSerLysSerLysG1 498
||| :||||| :|||
1038 TATGGGTAAATATT...AATGTTACTTTCACACCAAGT 1071
||| :||||| :|||
498 yAlaGlyAsnIleValAlaValSerPheSerGlnSer 510
||| :||||| :|||

```

seq_name: p1r2:B64635

seq_documentation_block:

toxin-like outer membrane protein HP0922 - *Helicobacter pylori* (strain 26695)
 C:Species: *Helicobacter pylori*
 C:Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999
 C:Accession: B64635
 R:Tom, J.F., White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalil, H.G.; Glodek, A.; McKenney, J.D.; Kelley, J.M.; Cotton, M.D.; Feldman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.; Nature 388, 539-547, 1997
 A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karik, P.D.; Smith, H.O.; Fraser, C.
 A:Title: The complete genome sequence of the gastric pathogen *Helicobacter pylori*.
 A:Reference number: A64520; M01D:97394467
 A:Accession: B64635
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-2529 <TOM>
 A:Cross-references: GB:AE000602; GB:AE000511; NID:g2314060; PIDN:AAD07969.1; PID:g231406

alignment_scores:

Quality: 108.00 Length: 374
 Ratio: 0.568 Gaps: 22
 Percent Similarity: 50.802 Percent Identity: 22.460

alignment_block:

US-09-839-894-9 x B64635

Align seg 1/1 to: B64635 from: 1 to: 2529

```

253 ACCAGTGAGAACTCTAGCAGTTCATCGGTACGGCTGACAAACAATATATAC 302
||||| :||||| :|||||
977 ThrSerAsnPheAsnAlaThrThrGlnLeuGlnLysThrAsnPheThr 993
||||| :||||| :|||||
303 ATTACAAATTTACGAAAAACAGTTTAATA..... 333
||| :||||| :|||
993 rLeu.....SerGlnSerLeuLeuAsnPheAsnGlyAspThr 1007
||| :||||| :|||
334 .....AAAAGAGCTACAAATTAAGGCTAATAACAATTAATG 372
||| :||||| :|||
1007 hrLeuGlnAsnAlaAsnAlaThrThrLeuGlnLysThrSerGlnAlaAla 1023
||| :||||| :|||
373 TTCAAA.....AGTGTAACTGCCCATCGGCTTAACACTT...AA 410
||||| :||||| :|||||
1024 PheLysAsnSerLeuThrLeuAspAsnAsnSerAsnLeuSerLeuAspAs 1040
||| :||||| :|||
411 CTCAGCTCATTTTAACGTATAATAAAACGGCGCT.....TCAGGTGCAA 454
||| :||||| :|||
1040 nGlnSerValLeuAsnAlaAsnAsnThrSerAlaPheAsnGlnAlaAs 1057
||| :||||| :|||
455 GTTATATATTAATATATCTCTCT...GGCGAATCAAAAAATTTGCTTTT 501
||||| :||||| :|||
1057 erLeuAsnIleTyrAsnGlySerGlnAlaThrPheAsnSerLeuPhePhe 1073
||| :||||| :|||
502 GGTGTATCTGGAGTCTCTCTGAGTTAAGACTAAAGACATATAG 551
||| :||||| :|||
1074 AsnGly.....GlyThrLeuSerLeuAsnAlaSerLysLysAs 1087
||| :||||| :|||
552 TGAACCTATGCACTTAACATATAATATCACTTAATTAATCTGAT. 600
||| :||||| :|||
1087 nAlaSerAsnAlaSerPheSerAsnAsnThrThrIleAsnLeuAspAs 1104
||| :||||| :|||
601 .....AAGCAAAATATTCAGATA 618
||| :||||| :|||
1104 erValLeuSerAlaSerAsnThrSerSerLeuAsnAlaAsnIle..... 1118
||| :||||| :|||
619 TGTATTACCTCAGTTCAAAAGTGACGCTCGCTGCATCTTAATCTGGCTC 668
||| :||||| :|||
1119 .....AsnPheGlnGlyAlaSerGlnAlaAspPhe..... 1128
||| :||||| :|||
669 AACTGTGGGGGACATATATTTGGAAGAAATCTGTTGATGTGCTTT 718
||| :||||| :|||
1129 ....GlyLysAsnThrIleIleAspThrAlaSerPheAsn.....Phe 1142
||| :||||| :|||
719 ATGATGATATAGTACTACAGCAGCTCTTGG.....GAGATA 756
||| :||||| :|||
1142 spSerAlaSerSerLeuAsnPheAsnAsnLeuThrAlaAsnGlyAlaLeu 1158
||| :||||| :|||
757 AGATTTCAGATTAACAATCTTAA..... 780
||| :||||| :|||
1159 AsnPheAsnGlyTyrThrProSerLeuThrLysAlaLeuMetSerValSe 1175
||| :||||| :|||
781 .....TCTGATGGGAAATTTATCTAAGAAATTA 811
||| :||||| :|||
1175 rGlyGlnPheValLeuGlnLysAsnGlyAspIleAsnLeuSerAspIleA 1192
||| :||||| :|||
812 AT.....GATGACACCAAGAAATTCATTAATCTGCTCACTCTC 852
||| :||||| :|||
1192 snIlePheAspAsnIleThrLysSerValThrTyr.....AsnIleLeu 1206
||| :||||| :|||
853 TTGGCGGGTAAAGTTTAATCTCAACAANGAAGCTCA..... 891
||| :||||| :|||
1207 AsnAlaGlnLysGlyIleThrGlyIleSerGlyAlaAsnGlyTyrGly 1223
||| :||||| :|||
892 .....TTAAATATGCTGACGCGAGCTCTCTCTGGAACAA 925
||| :||||| :|||
1223 sIleLeuPheTyrGlyLysIleGlnAsnAlaThrTyrSerAspAsnA 1240
||| :||||| :|||
926 AC.....TGGAATGAATTAACAGCTGTACACATGCCAAGATC 963
||| :||||| :|||
1240 snAsnIleGlnThrThrPheIleAsnProLeuAsnSerSerGlnIle 1256
||| :||||| :|||

```

964 AGTGTCCGGTGTG..... 978
1257 IleglnluserllelysasnlglyAspleuThrlleaglValleuAsnas 1273
979 TGTGGCCCTGGACGTTTGC 997
1273 nProAsnSerAlaSerAsnThrllePheAsnIleAlaProgluLeuTyra 1290
998 AATTGGATGCAAAAGTGAATAATCC..... 1023
1290 snTyrglnAlaSerlleysGlnAsnProThrglyTySerTyAspTySer 1306
1024 GAGCGTGACAAATAT..... ATGGTAATATTAATGTTACTTT 1061
1307 AspAsnclnAlaaglyThrTyTyTyTySerAsnIlelysglyleuPh 1323
1062 CACACCA..AGTAGTCAACA 1080
1323 eThrProlysglySerGlnThr 1330

seq_name: p1r2:A64904

seq_documentation_block:
Probable fimbrial protein b1502 - Escherichia coli
C.Species: Escherichia coli
C.Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 29-Sep-1999
C.Accession: A64904
R.Blatner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A.Title: The complete genome sequence of Escherichia coli K-12.
A.Reference number: A64720; MUID:97426617
A.Accession: A64904
A.Status: nucleic acid sequence not shown; translation not shown
A.Molecule type: DNA
A.Residues: 1-304 <BLAT>
A.Cross-references: GB:AE00247; GB:U00096; NID:g1787773; PIDN:AC74575.1; PID:g1787779;
A.Experimental source: strain K-12, substrain MG1655
C.Superfamily: fimbrial protein fimb
C.Keywords: fimbria

alignment_scores:
Quality: 107.00 Length: 357
Ratio: 0.629 Gaps: 19
Percent Similarity: 47.619 Percent Identity: 21.008

alignment_block:

US-09-839-894-9 x A64904 ..

Align seg 1/1 to: A64904 from: 1 to: 304

7 AAGATTTTATTT...ATTTTACATGTTGTTTCTCTACGACTTTTAC 53
|||||
8 LysValleuPhegylIleTyTyTyTyLeuMetAlaIleLysValAlPheAl 24
54 ATTGCTGATCGGAGATAAATTCGCCGAGATGAAGACATA..... 96
:|||||: :|||||
24 apheSerCysAsnValAsp.....glyTySerSerIleIleValag 38
97ACTAATATTTTGGCCCGCTGACGAAACGAATCTTCCCCAAA 141
|||||: :|||||
38 LyrThrIleSerValTyTyValAsnLeuAsp...ProValIleGlnProgly 53
142 CATTAATATA.....TTAAATACCATATTAACAGATACAGTGAAG 182
:|||||: :|||||
54 GlnAsnLeuValAlaAspLeuSerGlnIleIleSerCysTrpAsnAspTy 70
183 TCATACCTCTATGATGATGACTTTTATGTTTGTCTTCAC...A 229
70 rglglyTyTrpTyAsp.....ThrasPhisIleA 80

230 ATACACTTAATGAGCATGTCCACACGATGAGAAATCTACGACTTCATCG 279
||| :|||
80 snleuValglngly.....SerAla 86
280 GTACGGCGTGAACAATAATATACATTACATTACGAAAAAAGAGTTT 329
:|||||
87 PheAlagly..... 89
330 AATAAAAAGAGACTACAAATTAAGCTATAA...CAATTATGTGCA 376
90SerleuGlnSerTyTyTyTyTyTyTyTyTyTyTyTyTyTyTy 101
377 AAGGTGTAATCTGCCATCCGCTGACACTTAACACTTACGCTCATTTTAC 426
101 snAsnValThrTy 117
427 TGTAAATTAACCGCGCTTACGCTGCAAGTTTATTTATATTTCTGCG 476
118 IleglyAspTyThrProMetProleuProleuTyTyTyTyTyTyTyTyTyTy 132
477 TGGCGAAGTAAAAATTTGCTTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 526
133ThrProValGlyAlaAlaIleValIleValIleValIleValIle 144
527 AG.....TTAAGATAAAAAAGACATATAGTGAACCTAT 561
144 ysAlaIleGlyValIleAlaArgIleHisMetTyTyTyTyTyTyTyTyTyTy 160
562 GGAACT.....TACACTATAATATACATTTTAA..... 591
161 GlySerGlyAsnProArgAsnPheThrTyTyTyTyTyTyTyTyTyTyTyTyTy 177
592TTAAGTATTAAGGAAATATATACAG 616
177 nValValMetProThrglyGlyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTy 194
617 TATGTTTACCTCAGTTCAAAAGTACGCTGCGCTGATCTTAATCTTGCCT 666
194 alAspLeuProAspPheProglySerAlaIleIleProleu..... 207
667 CCAACTGTGTGGGGCACATATATTGGAAGAAATTCGTTGATGTCCTT 716
208GlyValTy 219
717 TTAATGATGATATAGTACATGACGAGCTCTTGGAGATAGATATGATCAG 766
219 eTy 235
767 ATAACAATCCTTAATCTGATGGGAAATTTATCTAAGAAATTAATGAT 816
235 snThrlaPro.....Asp 239
817 GACACCAAGAAATTCATATATCTTGTCTCTCTGCGGGGTAA... 864
240 AlaThrlYsAlaSerGlyValIleValIleThrlLeuMetArgAsnIleLys11 256
865AGTTAATCTCAACAATGGAACGTCAT 892
256 eLeuAlaThrlGlyAsnValSerleuGlyThrValAsnLysSerLysV 273
893 TAAATATGCTGACGAGCTTCTGTGAAACAAACTGATATAGATATTAAC 942
273 alProleuGlyLeuSerAlaThrlTyTyTyTyTyTyTyTyTyTyTyTyTyTyTy 289
943 GCTGTACATGCCAGAAATC 963
290 AlaIleTyThrlValGlnSerVal 296

seq_name: p1r2:C90892

seq_documentation_block:
Probable adhesin [imported] - Escherichia coli (strain O157:H7, substrain RIMD 050995
C.Species: Escherichia coli

C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
 C:Accession: C90892
 R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;
 gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
 DNA Res. 6, 11-22, 2001
 A:Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genc
 A:Reference number: A99629; MUID:21156231; PMID:11258796
 A:Accession: C90892
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-304 <HAY>
 A:Cross-references: GB:BA000007; PIDN:BA835530.1; PID:G13361573; GSPDB:GN00154
 A:Experimental source: strain O157:H7, substrain RMD 0509952
 C:Genetics:
 A:Gene: ECs2107
 C:Superfamily: fimbrial protein fimH

alignment_scores:
 Quality: 107.00 Length: 357
 Ratio: 0.629 Gaps: 19
 Percent Similarity: 47.619 Percent Identity: 21.008

alignment_block:
 US-09-839-894-9 x C90892 ..

Align seg 1/1 to: C90892 from: 1 to: 304

```

7  AAGATTTTATTT...ATTTTACATGTTTCTTCAGTACTTTTAC 53
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
8  LysValLeuPheGlyIleTyrLeuLeuMetAlaGlyLysValPheAl 24
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
54 ATTTGCTATCGGAGATTAATAATCCCGAGATGAAGCATA..... 96
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
24 APheserYsaInvalAsp.....GlyLysSerIleGlyAlaG 38
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
97 .....ACTAATATTTTGGCCCGCGTGACAGAACGATCTTCCCAA 141
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
38 LyrThrSerValTyrValAsnLeuAsp...ProValIleGlnProGly 53
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
142 CATATATATA.....TTAAATACCATATTACAGCATACGTAAG 182
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
54 GlnAsnLeuValAlaIAspLeuSerGlnHisIleSerCysTrpAsnAspTly 70
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
183 TCATCTCTGTATGATAGATGACTTTTATGTTGCTCTCTCAC...A 229
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
70 rGlyIlyTyrPyrAsp.....ThrsPhisIleA 80
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
230 ATACACTTAATGAGCATGTCCACACAGTGAATCTACAGATTCTACG 279
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
80 snLeuValGlnGly.....SerAla 86
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
280 GTCACGGGTGAACAATATAACATTACATTACGAAAAAGAGATT 329
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
87 PheAlaGly..... 89
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
330 AATAAAGAGAGCTACAAATTAAGGCTATRAA...CAATATTGTCTCA 376
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
90 .....SerLeuGlnSerTyrLysGlySerLeuTyrTyr 101
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
377 AAGAGTTTAAGTCCCATCCGCGCTAACACATTAACTGACATCTTAC 426
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
101 snAsnValThrTyrProPheProLeuThrAsnThrAsnValLeuAsp 117
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
427 TGTATAAAGACGGGCTTCAGGTGCAAGTTTATATATATATCTCCG 476
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
118 IleGlyAspLysThrProMetProLeuProLeuLysLeuTyrIle.... 132
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
477 TGGCAACTAAAAATTTGCTTTGGTGATATCTGGGATGCTACTCTGA 526
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
133 .....ThrProValGlyAlaIleAlaGlyLysValValIle 144
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
527 AG.....TTAAGAGTAAAAAGACGATATAGAGACCTAT 561
  
```

```

144 ysAlaGlyLysValIleAlaIleArgIleHisMetTyrLysIleAlaIleLeu 160
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
562 GGAAGT.....TACACTATTAATACATATTA.... 591
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
161 GlySerGlyAsnProArgAsnPheThrTrpAsnIleIleSerAsnAsn 177
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
592 .....TTACGATTAAGGAATATATTCAGA 616
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
177 rValValMetProThrGlyGlyCysThrValAspSerArgAsnValTrv 194
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
617 TATGTTACTCTCAAGTTCAAAAGTGACGCTCGCTGATTAACCTTGGCT 666
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
194 AlAsnLeuProAspPheProGlySerAlaGluIleProLeu..... 207
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
667 CCAACTGGTGGGGCCACATATATTGGAAGAAATTCGTGATATGCTT 716
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
208 .....GlyValTyrCysSer...SerGlnGlnLysLeuSerPh 219
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
717 TTATGATGATATAGTACTTAACAGCAGCTTTGGAGATTAAGATTACG 766
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
219 eryLeuSerGlyThrThrAspSerAlaArgIleVal...PheAla 235
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
767 ATACATCTCTAAATCTGATGGGAATTTTATCTAAGGAAATTAATGAT 816
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
235 snThrAlaPro.....Asp 239
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
817 GACACCAAGAAATTCATATACCTTTGTCACCTCTCTGGCGGTAA... 864
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
240 AlaThrLysAlaSerGlyValGlyValSerLeuMetArgAsnGlyLysI 256
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
865 .....AGTTTAACTCCCAACAATGAAGCGTCAT 892
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
256 eLeuAlaThrGlyGluAsnValSerLeuGlyThrValAsnLysSerLysV 273
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
273 alProLeuGlyLeuSerAlaThrTyrGlyGlnThrGlyAsnLysValSer 289
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
943 GCTGTACCATGCCAGAAATC 963
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
290 AlaGlyThrValGlnSerVal 296
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
seq_name: p1r2:F85725
seq_documentation_block:
probable adhesin, fimH type protein Z2206 [imported] - Escherichia coli (stra 57:
C:Species: Escherichia coli
C:date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: F85725
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
Ilier, L.; Grobeck, E.C.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: AB5480; MUID:21074935; PMID:11206551
A:Accession: F85725
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-304 <STO>
A:Cross-references: GB:AE005174; NID:G12515169; PIDN:AAG56266.1; GSPDB:GN00145; UWGP:
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: Z2206
C:Superfamily: fimbrial protein fimH

```

alignment_scores:
 Quality: 107.00 Length: 357
 Ratio: 0.629 Gaps: 19
 Percent Similarity: 47.619 Percent Identity: 21.008

alignment_block:
 US-09-839-894-9 x F85725 ..

Align seg 1/1 to: F85725 from: 1 to: 304

```

7  AAGATTTTATTT...ATTTTACATGTTTTCTCTCAGTACTTTTAC 53
  |||:|||||:| |||:| |||:|||||:| |||:| |||:|
8  LysValLeuPheGlyIleTyrLeuLeuMetValGlyValPheAl 24
  |||:|||||:| |||:| |||:|||||:| |||:| |||:|
54  ATTTGCTGTATCGGCAGATAAATTCGGAGATGAAGACATA..... 96
  |||:|||||:| |||:| |||:|||||:| |||:| |||:|
24  aPheSerCysAsnValAsp.....GlyGlySerSerIleGlyLag 38
  |||:|||||:| |||:| |||:|||||:| |||:| |||:|
97  ....ACTAATATTTTGGCCCGCTGACAGAAAGATCTTCCCCAAA 141
  |||:|||||:| |||:| |||:|||||:| |||:| |||:|
38  LyrThrSerValTyrValAsnLeuAsp...ProValIleGlnProGly 53
  |||:|||||:| |||:| |||:|||||:| |||:| |||:|
142  CATTAATATA.....TTAATAACCATATTTACAGATACATGAAG 182
  |||:|||||:| |||:| |||:|||||:| |||:| |||:|
54  GlnAsnLeuValValAspLeuSerGlnHisIleSerCysTrpAsnAsp 70
  |||:|||||:| |||:| |||:|||||:| |||:| |||:|
183  TCATACTCTGTATGATAGATGACTTTTATGTTGCTCTTCAC...A 229
  |||:|||||:| |||:| |||:|||||:| |||:| |||:|
70  rGlyGlyTrpTyrAsp.....ThraSphisIleA 80
  |||:|||||:| |||:| |||:|||||:| |||:| |||:|
230  ATACACTTAATGAGCATGTCACACAGAGAAATCCTAGCAGTTCATCG 279
  |||:|||||:| |||:| |||:|||||:| |||:| |||:|
80  snLeuValGlnGly.....SerAla 86
  |||:|||||:| |||:| |||:|||||:| |||:| |||:|
280  GTCAGCGGTAAACAATATACATTAACATTACGAAAAAAGACGTTT 329
  |||:|||||:| |||:| |||:|||||:| |||:| |||:|
87  PheAlaGly..... 89
  |||:|||||:| |||:| |||:|||||:| |||:| |||:|
330  AATTAATAAGAGAGCTACAAATTAAGGCTATAA...CAATTATGTCA 376
  |||:|||||:| |||:| |||:|||||:| |||:| |||:|
90  ....SerLeuGlnSerTyrTyrGlySerLeuTyrTrpA 101
  |||:|||||:| |||:| |||:|||||:| |||:| |||:|
377  AAGGTGTAATGCCCCATCCGGCTAACACTTAACAGCTCATTTAAC 426
  |||:|||||:| |||:| |||:|||||:| |||:| |||:|
101  snAsnValThrTyrProPheProLeuThrTrpAsnThrAsnValLeuAsp 117
  |||:|||||:| |||:| |||:|||||:| |||:| |||:|
427  TGTATATAAAACGGCGCTTCAGGTGCAAGTTTATATATATATTCCTCG 476
  |||:|||||:| |||:| |||:|||||:| |||:| |||:|
118  IleGlyAspLysThrPrometProLeuProLeuLysLeuTyrIle.... 132
  |||:|||||:| |||:| |||:|||||:| |||:| |||:|
477  TGGGGAAGCTAAAAAATTTGCTTTGGGTGATCTGGAGTCTACTCGA 526
  |||:|||||:| |||:| |||:|||||:| |||:| |||:|
133  ....ThrProValGlyAlaIleGlyValValIleL 144
  |||:|||||:| |||:| |||:|||||:| |||:| |||:|
527  AG.....TTAAGAGTAAAGAGCATATAGTGAGACCTAT 561
  |||:|||||:| |||:| |||:|||||:| |||:| |||:|
144  ysAlaGlyGluValIleAlaTrpGlieHisMetTyrLysIleAlaThrLeu 160
  |||:|||||:| |||:| |||:|||||:| |||:| |||:|
562  GGAAGCT.....TACATATAAATTCACATTAATAA..... 591
  |||:|||||:| |||:| |||:|||||:| |||:| |||:|
161  GlySerLysAsnProArgAsnPheThrTrpAsnIleIleSerAsnAsnSe 177
  |||:|||||:| |||:| |||:|||||:| |||:| |||:|
592  ....TTAAGTATAGAGAAATATTCAGA 616
  |||:|||||:| |||:| |||:|||||:| |||:| |||:|
177  rValValMetProThrGlyGlyCysThrValAspSerArgAsnValThrV 194
  |||:|||||:| |||:| |||:|||||:| |||:| |||:|
617  TATGCTTACTCAGTTCAAAAGAGACGCTGCGTCGATCTTAAGTTCGCT 666
  |||:|||||:| |||:| |||:|||||:| |||:| |||:|
194  alAsnLeuProAspPheProGlySerAlaGluIleProLeu..... 207
  |||:|||||:| |||:| |||:|||||:| |||:| |||:|
667  CCAAGCTGGGGGACATATATATATGAGAAATCTGTGATATGCTT 716
  |||:|||||:| |||:| |||:|||||:| |||:| |||:|
208  ....GlyValTyrCysSer...SerGlnGlnLysLeuSerPh 219
  |||:|||||:| |||:| |||:|||||:| |||:| |||:|
717  TTAATGATGATATAGTACTAACACAGCCTCTTGAGATAGATTTGAG 766
  |||:|||||:| |||:| |||:|||||:| |||:| |||:|
219  eTyrLeuSerGlyThrThrThrAspSerAlaArgGlnVal...PheAla 235
  |||:|||||:| |||:| |||:|||||:| |||:| |||:|
767  ATAACAATCTTAATCTGATGGAAATTTATCTAAGAAAAATTAATGAT 816
  |||:|||||:| |||:| |||:|||||:| |||:| |||:|

```

```

235  snThrAlaPro.....Asp 239
  |||:|||||:| |||:| |||:|||||:| |||:| |||:|
817  GACACCAAGAAATTCATATACCTTGTCTGACTTCTCTGGCGGTAA... 864
  |||:|||||:| |||:| |||:|||||:| |||:| |||:|
240  AlaThrLysAlaSerGlyValGlyValSerLeuMetLysAsnGlyLys 256
  |||:|||||:| |||:| |||:|||||:| |||:| |||:|
865  ....AGTTTAAGTCCAAATGGAAGCTCAT 892
  |||:|||||:| |||:| |||:|||||:| |||:| |||:|
256  eLeuAlaThrGlyGluAsnValSerLeuGlyThrValAsnLysSerLys 273
  |||:|||||:| |||:| |||:|||||:| |||:| |||:|
893  TAAATATGCTGACGACGCTTCTCTGAAACAAACGATGATAGATTA 942
  |||:|||||:| |||:| |||:|||||:| |||:| |||:|
273  alProLeuGlyLeuSerAlaThrTyrGlyGlnThrGlyAsnLysValSer 289
  |||:|||||:| |||:| |||:|||||:| |||:| |||:|
943  GCTGTACACATGCCGGAATC 963
  |||:|||||:| |||:| |||:|||||:| |||:| |||:|
290  AlaGlyThrValGlnSerVal 296
  |||:|||||:| |||:| |||:|||||:| |||:| |||:|

seq_name: p1r2:A38109

seq_documentation_block:
autolysin - Enterococcus faecalis
C:Species: Enterococcus faecalis
C:date: 24-Jul-1992 #sequence_revision 24-Jul-1992 #text_change 15-Oct-1999
C:Accession: A38109
R:Bellevue, C.; Potvin, C.; Trudel, J.; Asselin, A.; Bellemare, G.
J. Bacteriol. 173, 5619-5623, 1991
A:Title: Cloning, sequencing, and expression in Escherichia coli of a Streptococcus
A:Accession: A38109
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-671 <BEL>
A:Cross-references: GB:M58002; NID:9153658; PIDN:AAA67325.1; PID:9829194

alignment_scores:
Quality: 106.00 Length: 309
Ratio: 0.716 Gaps: 17
Percent Similarity: 47.896 Percent Identity: 23.625

alignment_block:
US-09-839-894-9 x A38109 ..
Align seg 1/1 to: A38109 from: 1 to: 671

121  AGAAGACAATCTTCCCAACAACATATATA...TTAATAACCATATTAC 167
  |||:|||||:| |||:| |||:|||||:| |||:| |||:|
309  ArgTyrAlaThrAspProSerTyrAsnAlaLysLeuAsnValIleTh 325
  |||:|||||:| |||:| |||:|||||:| |||:| |||:|
168  AGCATACAGTGAAGATCATCTCGTATGATAGATGACTTTTATGCT 217
  |||:|||||:| |||:| |||:|||||:| |||:| |||:|
325  rAlaTyr.....AsnLeuThrGlnTyrAspThrProSerSerGlyAla 340
  |||:|||||:| |||:| |||:|||||:| |||:| |||:|
218  TGTCTCTCACATACACTTAATGAGACATGTCACACAGTGAATCTCT 267
  |||:|||||:| |||:| |||:|||||:| |||:| |||:|
340  snThrGlyGlyGlyThrValAsnProGlyThrGlyGlySerAsnAsnGln 356
  |||:|||||:| |||:| |||:|||||:| |||:| |||:|
268  AGAGTTCATCGGTC.....AGCGGTAAACA.....AA 296
  |||:|||||:| |||:| |||:|||||:| |||:| |||:|
357  SerGlyThrAsnThrTyrTyrThrValLysSerGlySphisThrLeuAsn 373
  |||:|||||:| |||:| |||:|||||:| |||:| |||:|
297  TATTAACATTAACAATTT.....ACGAAAAAAGAAGTTTA.... 330
  |||:|||||:| |||:| |||:|||||:| |||:| |||:|
373  sIleAlaIleGlnTyrGlyValSerValAlaAsnLeuArgSerTrpAsn 390
  |||:|||||:| |||:| |||:|||||:| |||:| |||:|
331  ..ATAAAAAGAGACTACAAATTAAGGCTATAAACAATTAATTTGTTCAA 378
  |||:|||||:| |||:| |||:|||||:| |||:| |||:|
390  LysLeuSerGlyAspLeuIlePheValGlyGlnLysLeuIleValLysLys 406
  |||:|||||:| |||:| |||:|||||:| |||:| |||:|
379  AGGTTAAGTCCATCGGCGCTAACACTTAAGCTGACATCTTAACTG 428
  |||:|||||:| |||:| |||:|||||:| |||:| |||:|
407  GlyAla.....SerGlyAsnThrGlyGlySerGlyLysGlnGlyLys 420
  |||:|||||:| |||:| |||:|||||:| |||:| |||:|

```

429 TAATATAAAACGGCGCTTCAGTGCAGAATTATTTAATATATCCGCGC 478
||||| ||||| :||| :|||
420 rAsnAsn.. GlnSerIyThrsnThrYrYrThValLysSere 436
::: |||
429 GCGAACATAAAAAATTTGCCCTTTGGTGCTATCTGGGATGCTACTCTGAAG 528
|||
436 LysSprThrLeuAsn..... 440
529 TTAAAGCATAAAAGACGATATAGTAGACCTATGGAACCTTACTATAA 578
::: |||||
441LysIleAlaIaIntYrGly..... 447
579 TATCACTATTAAATTAACTGATGATGAAGGAATATTCGATATGTACTATC 628
::: ||||| :||| :|||
448 ValThrVal..... AlaAsnLeuArgSerTrp..... 456
629 AGTTCAAAGATGACGCTCGCGTCGATCTTAACCTG..... 663
::: ||||| :::
457AsnGlyIleSerGlyAspLeuIlePheValGlyIntLysLeu 470
664CGTCCAACCTGGTGGGGCACATATATTGGAAGAATATCTG 704
::: ||||| :|||
471 IleValLysLysGlyThrSerGlyAsnThr..... 480
705 TGATATGTGCTTTTATGATGATATAGTACTAACACAGCTCTTGGAGA 754
481GlyLysSerAsnGlyLysSer..... 488
755 TAGATTTTCAGTATACAATCCTAAATCTGATGGGAATTTTCTTACAG 804
489AsnAsnAsnGlnSerIeLyThrsnThrYrYrThrIle 501
::: ||||| :::
805 AAAATAAATGATGACACCAAAAGAAATGCA..... TATACTTGTCA.. 846
||| ::||| ::|||
502 LysSerGlyAspThrLeuAsnLysIleAlaIaIaIntYrGlyAlaSera 518
847CTTCTCTTGGCGG 859
518 lAlaAsnLeuArgSerTrpAsnGlyIleSerGlyAspLeuIlePheAlaG 535
860 GTAAGAGTTTAACCTCCACAACAAATGAGACGTCATTAATATGCTGACG 909
||| ::||| :||| :|||
535 LysIntLysIleIleValLysGlyThrSerGlyAsnThrGlyLysSer 551
910 GCCTTCTGTGGAACAACCTGGAATACA 936
552 SerAsnGlyLysSerAsnAsnAsnGln 560

```

seq_name: plr2:AC2507
seq_documentation_block:
hypothetical protein all71235 [imported] - Anabaena sp. (strain PCC 7120) plasmid pcc7120
C:Species: Anabaena sp.
A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002
C:Accession: AC2507
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kurlitz, T.; Sasamoto, S.; Watanabe, A.; Iriguchihashi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-Fixing Cyanobacterium Anabaena
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AC2507
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-843 <R>
A:Cross-references: GB:BA000020; PIDN:BAF78319.1; PID:g17135773; GSPDB:GN00180
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all71235
A:Genome: plasmid

```

Alignment_scores:	Quality:	105.50	Length:	376
Ratio:	0.567	Gaps:	18	
Percent Similarity:	49.468	Percent Identity:	18.883	
alignment_block:				
US-09-839-894-9 x AC2507 ..				
Align seg 1/1	to: AC2507	from: 1	to: 843	
229	AATACACTTAAATGGAGCATGTCACCAAGTGAGAAAT.....	264	
229	264	
297	AsnIleAspGlySerLeuValPheThrGlnAsnHisGlyPheIysThr	313	313	
265	313	
313	rcIyIaValIyIsIleAspIaGlnSerLeuAsnIleGlnIySerSer	330	330	
226	ATATACATTTACA.....	315	TTTACG.....	
330	snIeUaIeUeSerIaIeTyThrSerAsnPhcIyPheThrProGly	346	346	
316	GAATAAAGATTTATATATAAGAGACTCAATATTAAGCGTATAACA	365	365	
347	GIsertIleGlnLeuAspValIyAspValThrIleGlnIyGlnIle	363	363	
366	ATTATTTGTCAAAAGTGTAACTGCCATCCGGCCTA...ACACTTA	412	412	
363	ealThrThrThrPheThrAsnAlaProSerGlyLeuIleThrIleAsn	380	380	
413	CAGCTCATTTTAAC.....	438	TGTATATAAAC	
380	erAsnSerLeuIyIsIleSerGlyAspThrProSerTyTAlaAsnPro	366	366	
439	CGCGCTTAGCGGCAAGCTTATTTATTAATTCCT.....	474	474	
397	GlyLeuGlyGlyIleAsnThrPheSerTySerSerGlyIyGlyIyAs	413	413	
475	...GCTGGCGAACAATAAAATTTGCCCTTTGGT.....GGATCTGGG	514	514	
413	PhIeIaGlyIyIsIleAsnAsnIleIleGlyIeLeuAspGlyValPhe	430	430	
515	AT.....	537	GCTACTCTGAAGTTAAGCA	
430	snThrValAlaSerGlySerGlyIaGlyIyAsnLeuPheLeuGlyLeu	446	446	
538	AAAGACATATAGTAGAGACCTATGAGAACTTACACTTAATATTCATAT	587	587	
447	GUAAsnLeuIleIleIyAspGlyIyAlaSerLeuGlySerSerThrI	463	463	
588	TAAATTAAGTATAGGAGCAATATTCAGATGTGTACTCAGTTCAAA	637	637	
463	eArgSerGlyGlnGlyIyAsnVal.....	474	474	
638	GTAGCGTGGCGTCAGCTTAAC.....	672	TTGCGTCAACT	
474	ySerGlnAsnIleAspIleSerGlyGlnSerAlaLeuLeuArgPro	490	490	
673	695	GTCGGGCGCAATATTTGGAG	
491	AsnIleThrSerSerThrPheGlyIyAsnIyGlyIyAsnIleAspI	507	507	
696	AAATCTGTGATATAGTCTTTTATGATAGTATATAGTAAAGAGAGCT	745	745	
507	eAsnThrLeuAsnLeuIleIleSerAsnGlyGlyIyIleSerSerSer	524	524	
746	CTTTGGAG.....	753	753	
524	hrLeuSerAlaGlyIyAlaGlyIyAsnIleSerIleAsnSerSerAsn	540	540	
754	782	TC	
541	IleAsnValIyGlyThrAsnIleAsnSerAsnSerProPheIleAs	557	557	

```
783 TGATGGGAATTTATCTAAGAAAATAATGATGACCAAGAAATG 832
      ::::::::::|||  |||  :::::  :::::
557 nserSerAsnPheLeu.....ValAspProAsnLeuGlnLysLeu 572
      ::|
833 CATATACCTTGTCACTCTCTTGCGGGTAAAGT..... 867
      |||  |||  |||  |||  |||  |||  |||  |||
572 eutYrArgInProProLeuLeuIleGlyGlnIleGlyAsnIlePheLeu 588
      ::::::::::TTAACTCCACAAATGSAAGTCATTAAAT..... 897
868 ..... 897
589 AsnThrAspIleIleAsnIleSerAsnGlyGlyLeuIleAsnIleArgAs 605
      :::::::::::|||||  :::::
898 .....ATTGCTGACGACGCTCTCTGGAACAACTGGAATAGAATTA 940
      ::|  |||  |||  |||  |||  |||  |||  |||
605 ngluglyValAsnAspIleGlyAsnIleArgIleSerAlaAsnThrIleA 622
941 CAGCTGTCACCATGCCAGAAATCAGTGTCCGGTGTGTGGCT... 987
      ::  :::::  |||  |||  |||  |||  |||  |||
622 snIleAsnSerGlnGlyGluValAsnIleThrThrIleGlyGly 638
      ::  :::::  |||  |||  |||  |||  |||  |||
988 GGACGTTTGCATTGGATGCCAAAAGTGAAATCCCGAGCTGGACATA 1037
      |||  |||  |||  |||  |||  |||  |||  |||
639 GlyAsnIleIleLeuAsnSerArg.....AsnLeuPn 649
1038 TATGGTAATATTATGTTACTTTCACA 1065
      ::::::::::|||  :::::
649 eleuAsnAsnSerArgIleThrIleThr 658
```



```

151 TTAATAACCATATTACAGCATACGAGAAAGTCATCTGTATGATAG 200
|||||
51 LeuSnsSerTyrLeuThrAlaTyrAsnGlySerHisHisLeuTyrAspAr 67
201 GATGACCTTTTATGTGTCTCTCCACAAATACATTATGAGCATGTC 250
|||||
67 gmeSerHeuLeuGlySerSerGlnAsnThrLeuAsnGlyAlaCysP 84
251 CAACAGTGAGAAATCTAGCAGTTCATCGTCAGCGGTGAAACAAATATA 300
|||||
84 roSerSerAspAlaProGlyThrAlaThrIleAspGlyGluThrAsnIle 100
301 ACATTACATTTTACGGAAAAAGAAATTAAATAAAGACAGCTCAAT 350
|||||
101 ThrLeuGlnPheThrGluLysArgSerLeuIleLysArgGluLeuGlnI 117
351 TAAAGGCTATAACAAATATTGTTCAAAAAGTTAAAGTCCCATCCGCC 400
|||||
117 elySolYTrLysGlnPheLeuPheLysAsnAlaAsnGlyProSerLysL 134
401 TAAACCTTAACCTAGCTATTTAACCTTAATTAATAAAGCGGCTTCAG 450
|||||
134 euAlaLeuAsnSerSerHisPheGlnGlySAsnArgGluGlnAlaSerGly 150
451 GCAAGCTTATATTTATATATATTCCTGCTGCGCAACTAAAAATTTGCC 500
|||||
151 AlaThrLeuSerLeuTyrIleProAlaGlyGluLeuAsnLysLeuProPh 167
501 TGGTGATCTCGGATGCTACTCTCAACTTAAGATAAAGACATATA 550
|||||
167 egLYGlyValTrpAsnAlaValLeuLysLeuAsnValLysArgAlaGlyTyr 184
551 GTGAGACCTATGGAACCTTACACTATATATACATTAAATTAATTA 600
:::
184 splrThrTyrGlyThrTyrThrIleAsnIleThrValAsnLeuThrSp 200
601 AAGGAAATATTACAGATATGTTACCTCAGTTCAAAAGTGACGCTCGCT 650
|||||
201 LysGlyAsnIleGlnIleTrpLeuProGlnPheLysSerAsnAlaArgVa 217
651 CGACTTACTTGGTCCCACTGGTGGGGGACATATTTGGAGAATTT 700
|||||
217 lAsPLeuAsnLeuArgProThrGlyGlyThrTyrIleGlyAlaGAsnS 234
701 CTGTGATATGCTCTTTATGATGATATAGTACTACAGCAGCTCTTG 750
|||||
234 ervAlaSpmecCysPheTyrAspGlyTyrSerThrAsnSerSerLeu 250
751 GAGATAGATTTACAGATTAACAATCTTAATCTGATGGAAATTTATCT 800
|||||
251 GluIleArgPheGlnAspAsnSerLysSerAspGlyLysPheTyrLe 267
801 AAGGAAATTAATATGATGACACCAAAAGAAATTCATATCTTGTACCTC 850
|||||
267 ulYSysLleAsnAspAspSerLysGluLeuValTyrThrLeuSerLeuL 284
851 TCTTGGCGGTAAAGTTAACTCCACAAATGGAACGCTATTAAATATT 900
|||||
284 euLeuAlaGlyLysAsnLeuThrProThrAsnGlyGlnAlaLeuAsnIle 300
901 GCTGACGACGCTCTCTGGAACAACTGGAATAGATATACAGCTGTGAC 950
|||||
301 ..AsnThrAlaSerLeuGluThrAsnTrpAsnArgIleThrAlaValTh 316
951 CATCCGCAATTCAGTCTTCGGGTGTGTGCTGGCTGGAGCTTTGCAT 1000
|||||
316 rmeProGluIleSerValProValLeuGlyStrProGlyArgLeuGlnL 333
1001 TGGATGCAAAAGTGAATAATCCGAGGCTGGAATATATGGAATATT 1050
|||||
333 euAspAlaLysValLysAsnProGluAlaGlyGlnTyrMeGlyLysnIle 349
1051 AATGTTACTTTCACACCAAGTAGTCAAAACACTC 1083

```

```

seq_name: SwissProt_40:SAGL_YEAST
350 LysIleThrPheThrProSerSerGlnThrLeu 360
:::
seq_documentation_block:
ID SAGL_YEAST STANDARD; PRT; 650 AA.
AC P20840;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Alpha-agglutinin precursor (Ag-alpha-1).
GN SAGL OR AGAL1 OR YUR004C OR J1418.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90014768; PubMed=2677666;
RA Lipke P.N., Wojciechowicz D., Kurjan J.;
RT "Ag alpha 1 is the structural gene for the Saccharomyces cerevisiae
RT alpha-agglutinin, a cell surface glycoprotein involved in cell-cell
RT interactions during mating."
RL Mol. Cell. Biol. 9:3155-3165(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=S288C;
RC MEDLINE=90005993; PubMed=2676603;
RA Hauser K., Tanner W.;
RT "Purification of the inducible alpha-agglutinin of S. cerevisiae and
RT molecular cloning of the gene."
RL FEBS Lett. 255:290-294(1989).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / FY1679;
RA de Haan M., Smit P.H.M., Grivell L.A.;
RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
RN [4]
RP PARTIAL SEQUENCE, CARBOHYDRATE-LINKAGE SITES, AND DISULFIDE BONDS.
RX MEDLINE=96064684; PubMed=7592821;
RA Chen M.-H., Shen Z.-M., Bobin S., Kahn P.C., Lipke P.N.;
RT "Structure of Saccharomyces cerevisiae alpha-agglutinin. Evidence for
RT a yeast cell wall protein with multiple immunoglobulin-like domains
RT with atypical disulfides."
RL J. Biol. Chem. 270:26168-26177(1995).
CC -|- FUNCTION: CELL SURFACE GLYCOPROTEIN PROMOTING CELL-CELL CONTACT
CC TO FACILITATE MATING. (SACCHAROMYCES CEREVISIAE A AND ALPHA CELLS
CC EXPRESS THE COMPLEMENTARY CELL SURFACE GLYCOPROTEINS A-AGGLUTININ
CC AND ALPHA-, RESPECTIVELY, WHICH INTERACT WITH ONE ANOTHER TO
CC PROMOTE CELLULAR AGGREGATION DURING MATING.
CC -|- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC (POSSIBLE).
CC -|- INDUCTION: BY EXPOSITION TO PHEROMONE (A-FACTOR) SECRETED BY THE
CC OPPOSITE MATING TYPE CELLS (TYPE A).
CC -|- PTM: N-GLYCOSYLATED AND O-GLYCOSYLATED.
CC -|- SIMILARITY: TO C. ALBICANS ALS1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M28164; AAA34417.1;
DR EMBL; X16861; CAA34752.1;
DR EMBL; X87611; CAA60926.1;
DR EMBL; Z49504; CAA89526.1;
DR PIR; S22835; S22835.
DR SGD; S0003764; SAGL.
KW Glycoprotein; Cell adhesion; Signal; GPI-anchor; Repeat.

```

```

alignment_scores:
    Quality: 114.00      Length: 469
    Ratio:   0.573       Gaps:   23
    Percent Similarity: 42.431     Percent Identity: 20.896

alignment_block:
US-09-839-894-9 x SAGI_YEAST ..

Align seg 1/1 to: SAGI_YEAST from: 1 to: 650

10 ATTTTATTTATTTTTACATTGTTTTCCTCAGACTT.....48
|||||:|||||:|||||:|||||:|||||:|||||:
8 IletenrtpneuseserleualaleualaSerallaileasnlleasnns 24
49 .TTTACATTTGGCT.....GTATCGGCAGATAAATATC 79
:::|||||||:|||||:|||||:|||||:|||||:|
24 pletmhpnesersnleugluiletmProeulethalasnllyslmp 41
80 CCGCA.....GATGAAGCATATAATATTTT 108
|| || |||||:|||||:|||||:|||||:
41 rokspsngllytptthralatrhpnasppheserllelaaslaspaser 57
109 GGCCCGCGTACAGAACGA.....TCTTCCCC..AA 140
::: |||:|:|:|:|:|:| | | | | | :

```

[illegible]

816 TGACACCAAGAAAT.....GCATPACTTTGTCA..... 846
 342 raspleuThrSerIleAsnThrSerIleThrSerIleSerIleSerIle 359
 847CTTCTC 852
 359 hValGluThrGlyAsnArgThrThrSerGluValIleSerHisValVal 375
 853 TTGGCGGGTAAAGTTTAACTCCACAAATGCAAGCTCAATAATATTCG 902
 376 ThrThrSerThrIleuSerProThrIleThrSerLeuThrIleal 392
 903 TGACGACGCTTCTCTGAAACAAATCGAAT..... 933
 392 aglnThrSerIleTyrSerThrIleSerHisIleThrValGlyThrAspI 409
 934AGAATP 939
 409 leHisThrThrSerGluValIleSerAspValGluThrIleSerArgIu 425
 940 ACAGCTGTACCATGCCAGAAATCAGTCCGCTGTGTGTCGCTCG 989
 426 ThrIleSerThrValValAlaAlaProThrSerThrGlyThrPhrG1 442
 990 ACCTTTG 996
 442 yAlaMet 444

seq_name: SwissProt_40:GAS3_YEAST

seq_documentation_block:

ID GAS3_YEAST STANDARD; PRT; 524 AA.

AC 003655; (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE GAS3 protein precursor.
 GN GAS3 OR YMR215W OR YMR261.09.
 OS Saccharomyces cerevisiae (Baker's yeast).
 CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 CC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / AB972;
 RA Dedman K., Brown D., Bowman S., Barrell B.G., Rajandream M.A.,
 RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP IDENTIFICATION.
 RX MEDLINE=20529944; PubMed=11079560;
 RA Pardo M., Ward M., Bains S., Molina M., Blackstock W., Gill C.,
 RA Nombela C.;
 RT "A proteomic approach for the study of Saccharomyces cerevisiae cell
 wall biogenesis";
 RL Electrophoresis 21:3396-3410(2000).
 CC -I- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
 (Potential).
 CC -I- SIMILARITY: BELONGS TO THE GAS1 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Z49809; CAAB9930.1; -;
 DR COMPLETFAST-2DPAGE; Q03655; -;
 DR SGD; S0004828; GAS3.
 KW Glycoprotein; Membrane; GPI-anchor; Signal.

FT SIGNAL 1 21 POTENTIAL.
 FT CHAIN 22 524 GAS3 PROTEIN.
 FT PROPEP 2 524 REMOVED IN MATURE FORM.
 FT DOMAIN 458 496 SER-RICH.
 FT CARBOHYD 201 201 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 269 269 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 350 350 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 385 385 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 404 404 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 422 422 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 504 504 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 524 AA; 56793 MW; 3A2BED0BD3ED8690 CRC64;

alignment_scores:
 Quality: 108.00 Length: 279
 Ratio: 0.735 Gaps: 18
 Percent Similarity: 52.688 Percent Identity: 24.731

alignment_block:
 US-09-839-894-9 x GAS3_YEAST

Align seg 1/1 to: GAS3_YEAST from: 1 to: 524

355 GGCTATAACAATTA.....FTGTCAAAAGTGTAACTGCCCGCGG 398
 264 GLYTYRAspLysLeuAsnSerThrPheGluAspAlaValIlePro..... 278
 399 CCTAACCTTAACCTCAGCTCATTTTAACTGTAATAAAACGCCGCT.... 444
 279LeuIlePheSerGluTyrGlyCysAsnLysAsnThrProArgT 293
 445TCAGTGAAGTTTATATTTATATTCCTCTGCGCACTAATAA 489
 293 hrpPheAspGluValSerGluGlyLeuTyr.....GlyGlyLeuLys 306
 490 AATTTCCTTTTGTGTGATCTGATCTGATCTGAAGTAAAGTAA 539
 307 AsnValPheSerGlyGlyLeu.....Va 314
 540 AAGACATATAGTAGAGCTTAGGAACTTACACTATATAATCATTATTA 589
 314 IYrGluTyrThrGluGluAlaLysAsnTyrGlyLeu.....ValL 328
 590 AATTACTGATAGGAATATACATATATGTTACTCTAGTCAAAAGT 639
 328 yLeuAspAspSerGlySerLeu.....ThrTyrLysAsp 339
 640 GAC...GCTCGCGTCGATCTTAAGTTCGT.....CCAACGTG 674
 340 AspPheValAsnLeuGlnLeuLysAsnValSerLeuProThrTh 356
 675 TGGGGCACATATATTGGAAGAAATTCGTGTGATATGTCG..... 714
 356 rLysGluSerGluIleSerSerAspSerIleTyrLysCysAspAsnSera 373
 715TTTATGATGATATATAGTACTAGACGAGCTCTTGAG 753
 373 lalleThrAsnIleTyrSerGlyPheGlyThrAsnSnbPheThrLeuPro 389
 754ATAAGATTTCAGGATTAACAATCC 776
 390 SerGlnProAlaGluIleAlaAsnMetIleGluTyrGlyValAsnLyn 406
 777 TAAATGTGATGGAA..... 792
 406 rAsnThr...GlyLysIleLeuThrAspTyrAlaValProThrThPhea 422
 793 ...TTTATCTAAGGAATAATAGATGACACCAAAATAATGCAATFACT 840
 422 snTyrThrIleLysAsnLysAspAspThr.....IleSerAlaThr 436
 841 TTGTCACTTCTTGGCGGGTAAAGTTTAACTCAACAATGGAACGTC 890

```

437 IleserTyrAspLysAla...AsnSerLeuAsnGluLeuAspValThrAl 452
891 ATTAATATTCGTGACGAGCTTCTCTGGAACAACAACGATGATA...A 937
452 athThrValAlaLysSerAlaSerThrSerGlnSerSerSerThr 469
938 TTACAGCTGTCACCATGCCAGAAATCAAGTGTCCGGTGTGTGGCT 987
469 eutThrSerSerThrSerProSerSerSerThrGlySerSerSerThr 485
988 GGACGTTTGCAATTGATGCAAAAGTGGAAAATCCGAGCGTGACATA 1037
486 Gly.....SerSerSerAlaSerSerSerSerLysSerLysG 498
1038 TATGGGTATAT...AATGTTACTTTCACACCACT 1071
498 yAlaGlyAsnIleValAsnValSerPheSerGlnSer 510

seq_name: SwissProt_40:YDEQ_ECOLI
seq_documentation_block:
ID YDEQ_ECOLI STANDARD; PRT; 304 AA.
AC P77588;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical fimbrial-like protein ydeq precursor.
GN YDEQ OR B1502.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=97426517; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE=97251357; PubMed=9097039;
RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakabe S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saito N.,
RA Sempel G., Seki Y., Sivasubram S., Tagami H., Takeda J.,
RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horikuchi T.;
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 28.0-40.1 min region on the linkage map.";
RL DNA Res. 3:363-377(1996).
CC -1- SIMILARITY: TO TYPE-1 FIMBRIAL SUBUNITS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL: AE000247; AAC74575.1; -
CC EMBL: D90793; BAA15183.1; -
CC EMBL: D90793; BAA15183.1; -
CC EcoGene: Egl3799; ydeq.
CC InterPro: IPR000259; Fimbrial.
CC Pfam: PF00419; Fimbrial; 1.
CC Hypothetical protein; Fimbria; signal; Complete proteome.

```

```

FT SIGNAL 1 26 POTENTIAL.
FT CHAIN 27 304 HYPOTHETICAL FIMBRIAL-LIKE PROTEIN YDEQ.
SQ SEQUENCE 304 AA; 32069 MM; 8153C8653087D99A CRC64;

alignment_scores:
Quality: 107.00 Length: 357
Ratio: 0.629 Gaps: 19
Percent Similarity: 47.619 Percent Identity: 21.008

alignment_block:
US-09-839-894-9 x YDEQ_ECOLI ..
Align seg 1/1 to: YDEQ_ECOLI from: 1 to: 304

7 AAGATTTTATTT...ATTTTACATTTGTTTCTCTGATGACTTTTAC 53
|||||
8 LysValIleuPheGlyIleTyrLeuLeuLeuMetAlaGlyLysValPheAl 24
54 ATTTGCTGTATCGGAGATAAATTCGCCGAGATGAAGACATA..... 96
|||||
24 aPheSerCysAsnValAsp.....GlyGlySerSerIleGlyAlaG 38
97 .....ACTAATATTTTGGCCCGCGTGACAGACGAATCTTCCCCCAA 141
|||||
38 LyrThrThrSerValTyrValAsnLeuAsp...ProValIleGlnProGly 53
142 CATATATATA.....TTAAATACCATATTTACAGCATACAGTGAAG 182
|||||
54 GlnAsnLeuValAlaLysPheSerGlnHisIleSerCysTyrPAsnAspTly 70
183 TCATACCTGTATGATAGATGACTTTTATGTTGTTGCTCTGCAC...A 229
|||||
70 rGlyGlyTyrPtyrAsp.....ThrAspHisIleA 80
230 ATACACTTAATGAGCATGTCACCAACAGTGAGAAATCTACGATTCATCG 279
|||||
80 snLeuValGlnGly.....SerAla 86
280 GTCAGCGGTGAACCAATATATACATTACATTTACGAAAAAAGAGTTT 329
|||||
87 PheAlaGly..... 89
330 AATAAAGAGAGCTACAAATTAAGCCTATATA...CAATTATTTGTCA 376
|||||
90 .....SerLeuGlnSerTyrLysGlySerLeuTyrTPA 101
377 AAGAGTTTAACTGCCATCCGCGCTTACACTTACCTGATGATTTTAC 426
|||||
101 snAsnValThrTyrProPheProLeuThrThrAsnThrAsnValLeuAsp 117
427 TGTATATAAAGCGCGCTTACAGTGAAGTTATATTTATTTATTTCTGCG 476
|||||
118 IleGlyAspLysThrProMetProLeuProLeuLysLeuTyrIle.... 132
477 TGGCGAAGTAAATAATTTGCTTTTGGTGATCTGGAGATGCTACTCGA 526
|||||
133 .....ThrProValGlyAlaAlaGlyLysValIleAl 144
527 AG.....TTAAGAGTAAAGACGATATAGGAGCCAT 561
|||||
144 ysaIaGlyGluValIleAlaArgIleHisMetTyrLysIleAlaThrLeu 160
562 GGAAGT.....TACACTATAAATACACTATTATAA.... 591
|||||
161 GlySerGlyAsnProArgAsnPheThrTyrPAsnIleIleSerAsnAsn 177
592 .....TTACTGATARAAGGAATATTCGA 616
177 nvaIValMetProThrGlyGlyCysThrValAspSerArgAsnValThrV 194
617 TATGTTTACCTGATCAAAAGTGACGCTGCGATCACTTAACCTTGGGT 666
|||||

```


660 CTTCGCTCACTGTTGGGGCACATAT.....ATTGAA 694
+ |||||::: |||
930 nlusserAlathrGlyAsnphethrleuGlnValAlaAspLysThrGly 947
695 GA...AATCTGTGATATGCTTTATAGTGAATTAAGTAAACG 741
+ |||
947 lUpRoAsnHsAsnGlnLeuThrLeuPheAspAlaSerAsnAlaThrArg 963
742 AGCTCTTGAGATAGATTCAGATTAACAAT...CCTAACTGATG 788
+ |||||::: |||
964 AsnAsnLeuGlnValThrLeuAlaAsnGlySerValAspArgGlyAlaTr 980
789 GAAATTTATCTAAGCAAAATTAATGATGACCAAGAAATTCATATA 838
+ |||||::: |||
980 pLysTyLysLeuArgAsnValAsnGly....ArgTyAspLeuTyra 995
839 CTTCGCTCACTGCTTGGCGGTAAGTTAACTCCACAATGAGACG 888
+ |||||::: |||
995 snProGlnValGlnLysArgAsnGlnThrValAspThrTrpAsnIleThr 1011
889 TCATTAAT.....ATTGCTGACGACCTCTCTGGAACAACATGGA 932
+ |||||::: |||
1012 ThrProAsnSpiIleGlnAlaAspAlaProSerAlaGlnInsAsnAsnGln 1028
933 TAGAATTACAGCTGTACACATGCCA 957
+ |||||::: |||
1028 uGlnIleAlaArgValGlnThrPro 1036

seq_name: SwissProt_40:YD93_METJA

seq_documentation_block:

ID YD93_METJA STANDARD; PRT; 608 AA.

AC 058788;

DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Hypothetical protein MJ1393.

GN MJ1393.

OS Methanococcus jannaschii.

OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;

OC Methanococcus;

OX NCBI_TaxID=2190;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;

RX MEDLINE=96337999; PubMed=8688087;

RA Bult C.J., White O., Olsen G.V., Zhou L., Fleischmann R.D.,

RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,

RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,

RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,

RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,

RA Uitterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,

RA Cotton M.D., Roberts K.M., Hursb M.A., Kaine B.P., Borodovsky M.,

RA Kleen H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;

RA "Complete genome sequence of the methanogenic archaeon, Methanococcus

jannaschii."

RT Science 273:1058-1073(1996).

RL -1- SIMILARITY: TO M.JANNASCHII MJ1394 AND A.FULGIDUS AF2028.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: U67579; AAB99403.1; -.

DR TIGR: MJ1393; -.

KW Hypothetical protein; Transmembrane; Complete proteome.

FT TRANSMEM 4 24 POTENTIAL.

SQ SEQUENCE 608 AA; 66768 MW; 010FAF1C29F8C73C CRC64;

alignment_scores:

Quality:	98.50	Length:	419
Ratio:	0.505	Gaps:	25
Percent Similarity:	46.539	Percent Identity:	20.286

alignment block:

US-09-839-894-9 x YD93_METJA

Align seg 1/1 to: YD93_METJA from: 1 to: 608

1	ATGAATTAAGATTTATTTT.....ATTGTTACATTTGTTCTCTTC	41
1	MetArgLysLeuIlePheMetAlaLeuLeuMetSerLeuPheIleG1	17
42	AGTACTTTTACATTGCTGTATCGGACGATTAATAATCC.....	81
17	YThrValPheGlyTyR.....GlyAspAsnProLeuTyVala	31
82GGAGATGAAGC	93
31	LatTyTyGlnLysTyRAsnIleThrGlyAsnThrThrGlyAspGlyLeu	47
94	ATACTAATATTTTGGCCCGCGTGACAGCAAGAACTTCCCAACA	143
48	ValSerSerThrIle.....GlutSerIleThrGlyTy	58
144	TAATATATTAATTAACCATATTACAGCATACAGTGAAGATCATCTGT	193
58	rIleValIleAsnAsn.....ThrGlyThrThrIleAsnAspThrLeu	73
194	ATGATAGGATGACCTTTTATGTTGTTGCTCTCACAATACA.....	234
73	YrAspValTrpValAlaValAsnIleSerAsnAsnIleThrGlyProGln	89
235CTTAATGA.....GCATGTC	251
90	ValTyValAsnGlyThrProLysGlyValPheIleGluSerSerAlaPr	106
252	AACCACTGAGAAAT...CTTAGCATTCATCG.....G	280
106	oAlaTyTrpAsnLeuProAsnAlaAsnThrTyRleHisIleProIleL	123
281	TCAGCGGTGAACCAATATTAACATTTACGTAATTAAGCAAGCTTA	330
123	eupRoAsnAsnSerTyRValIleIleLysPheAlaIleAspLysSerIle	139
331	ATAAAAAGAGAGCTACAAATTAAGGCTATTAACATTAATGTTCAAG	380
140	ThrGlyValProLeuIleIleAsnGln.....ThyTrSerAs	152
381	TGTTAACTGCCCATCGGCGCTAACACTTAAC...TCAGCATTTTA	427
152	pThrIleLysIleProSerGlnArgLeuSerAsnTrpSerValTyRLeuAsnI	169
428	GTAATTAACCGCGGCTTCAGTGCAAGTTATATTATATTCCTGCT	477
169	IleSerArgAsnValSerAla.....LeuProAla	178
478	GGCGAA.....CTAAAAAATTTGCC	497
179	ThrAspThrProValSerValIleMetThrLysTyRLeuSerAsnSpr	195
498	TTTTGGGTATCTGGATCTCTCTGGAAGTTAAGAGTAAAAAGCAG	547
195	O.....	195
548	ATAGTGAGCACTGAGCACTTACACT.....ATAATATCACTATT	588
196AsnAsnTyRGlYSerAspThrTrpAsnPheLeuAsnIleThrGly	210
589	AAATTAAGTATTAAGCAATATTAATTCAGATATG.....TT	623


```

211 AlaIlealaasnuglySerIleThrLeuTPspGlyProTyRphLe 227
624 ACCTCAGTTCAAA..... 636
227 uproGlyTyRasnspSerLeuThrTPrThGlyValIleasnThr 244
637 ..AGTACGCGTCGCGTCATCTTAATGTCGCAACTGTCGGGCGACA 684
244 hrlYsAsnAlaThrIleThrIleAsnIle.....ThGlyYasnAsnThr 258
685 TATATGAGAGAATTCGTGTGATATGCTTTATGATGATATATGAC 734
259 TyrThrAsnArgThnGlyThrLeuMetLysTyR.....GlyPheAlaVa 273
735 TAACAGCAGCTCTTGAGATATAGATTCAGATACATCTTAATCTG 784
273 .....IlePhePheGluPheAsnGlyThrLysSerG 284
785 ATGGCAATTTTATCTAAGAAATAATGATGACCAAGCAAAATGCA 834
284 LyrThrLysIle.....GluGlyIle 290
835 TATACT.....TTGTCACTCTCTTGCGGCTAAAGTTT 869
291 TyrAlaThrGlyTyRgLyIyAlSerAlaThrLysGluGlyProPhe 307
870 AACGCCAACAAATGA.....ACGTCATTAATATG 901
307 uAsnAlaSerSerGlyTyRgLyIleTPrTyGluSerAlaAsnValS 324
902 CTGACGCGCTCTCTGTAACAACAACTGAATATGATACGCTGCACC 951
324 eAsnLysAlaSer.....SerTyTyRAsnLeuThrHisValThr 338
952 ATGCCAGAAATC.....AGTGTCCGGTGTGTGTCGCGCTGACGTTT 995
339 IleTPrAlaValAsnGlySerAsnProValIleLeuAsnProPheAsn 355
996 GCAATTG 1002
355 eThrLeu 357

seq_name: SwissProt_40:SP21_YEAST

seq_documentation_block:
ID SP21_YEAST STANDARD; PRT; 758 AA.
AC P35209;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE SPT21 protein.
GN SPT21 OR YMR179W OR YMR8010.09.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94186069; PubMed=8138180;
RA Natsoulis G., Winston F., Boeke J.D.;
RT "The SPT10 and SPT21 genes of Saccharomyces cerevisiae.";
RL Genetics 136:93-105(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Churcher C.M., Barrall B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: REQUIRED FOR NORMAL TRANSCRIPTION AT A NUMBER OF LOCI IN
CC YEAST.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.1sb-sib.ch/announce/
CC or send an email to license@1sb-sib.ch).
CC -----
DR EMBL: L24436; AAA35078.1; -
DR EMBL: 249808; CAA89912.1; -
DR PIR: S47866; S47866.
DR SGD: S0004791; SPT21.
FT DOMAIN 127 144 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 672 682 ASP/GLU-RICH (ACIDIC).
SQ SEQUENCE 758 AA; 84697 MW; 7DB3FCFEE96705 CRC64;

alignment_scores:
Quality: 98.00 Length: 278
Ratio: 0.766 Gaps: 12
Percent Similarity: 46.043 Percent Identity: 21.942

alignment_block:
US-09-839-894-9 x SP21_YEAST ..

Align seg 1/1 to: SP21_YEAST from: 1 to: 758

124 AACGATCTCCGCCAACATATATATTAATTAACCATATTAACAGACATA 173
||| ||||| ||| .....||| ||||| |||
166 AsnIleSerAsnLysSerGlyArgValValAsnAsnGlnIle..... 179
174 CAGTGAAGTCATCTGTATATAGATGACCTTT..... 210
||| ||||| ||| .....|||
180 ...ProGluGluThrLeuGluValLysLeuArgPheThrLysValIleT 195
211 ....TTATGTTGCTCTTCTCAATACCTTAATGAGACATGTCACACC 255
||| ||||| ||| .....||| |||||
195 hAsnLeuArgThrSerGlyAsnAsnThrThrAsnSerArgLysSerCys 211
256 ACTGGAATCTCAGACGATTCATCGTCGTCACGCGTGAAACAAATATACAT 305
||| ||||| ||| .....||| |||||
212 LeuGlnMetProSerSerLeu.....ProSerAlaThrIle 223
306 ACAATTTACGGAAGAAAGATTATATAAAGAGAGCTACAAATTAAG 355
||| ||||| ||| .....||| |||||
223 uproPheThrProLysSerGlnSerLeuPheLysThrAsnGlnIleLys. 239
356 GCTATTAACAATTAATGTTCAAAAGTGAAGTACGCCACCGCCCTACA 405
||| ||||| ||| .....||| |||||
240 .....AsnSerArgAsnAlaArgThrThrIleThr 249
406 CTTACTCAGCTCATTTTAACGTATATAAAGCGGCTTCAGTGCAG 455
||| ||||| ||| .....||| |||||
250 IleAsn.....AsnThrAsnSerGlyThr.. 257
456 TTTATATTTATATATCTCTGCTGCGGAACATAAAATTTGCTTTGGTG 505
||| ||||| ||| .....||| |||||
258 .....ValGlyArgGlyThrAsnPromet.... 266
506 GTATCTGGATCTACTCTGAAGTTAAGAGTAAAGAGCATATAGTGA 555
||| ||||| ||| .....||| |||||
267 .....ProAlaProLysAlaValArgThrGln..... 275
556 ACCTATGGAACCTACATATAATATCACTATTAATTAAGTATAAGG 605
275 ..... 275
606 AATATTCAGATATGTTACTCAGTTCAAAAGTGACGCTCGCTGCATC 655
||| ||||| ||| .....||| |||||
276 ..SerLeuProIleTrp..... 280
656 TTAATCTGCTCCACTGCTGGGGGACATATATGGAAGAAATCTGT 705
||| ||||| ||| .....||| |||||
281 ..AsnLeuLysProAsnIleAlaAsnThrGlyPheProArgSerIle 296
706 GATATGTGCTTTAT.....GATGATATATAGTACTAA 737

```



```

196 snasprhrilaglyLeuphehisHis.....AsnglyglyThr 208
865 AGTTAACTCCACA...AATGAAACGTCATTAAATATGTCAGCAGC 911
209 AspheSerThrThrGluSerGlyIleTyrLysLysLeuTyrAspLeuAl 225
912 TTCTCTGGAACAACACTGATATAGATTATACGTCACCATGCCAGAA 961
225 aAspIleAsnGlnAsnAsnAsnThrIleAspSerTyrLeuLysGluSerI 242
962 TCACGTGTCGGGTGTGTGGCTGACGCTTGCAATTGATGACAAA 1011
242 LegInLeuTyrLeuAsnLeuGlyValAspGlyIleArgPheAspAlaVal 258
1012 CTGGAAATCCCGAGCTGACAA.....TATATGGGTATAT..... 1050
259 LysHisMetProGlnGlyTyrGlnLysSerTyrValSerIleTyrSe 275
1051 .....AATGTTACTTTCACA 1065
275 rSerAlaAsnProValPheThr 282

seq_name: SwissProt_40:RF1M_YEAST

seq_documentation_block:
ID RF1M_YEAST STANDARD: PRT; 413 AA.
AC P30775:
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Peptide chain release factor 1, mitochondrial precursor (MRF-1).
GN MRF1 OR YGL143C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93117110; PubMed=1475194;
RA Pel H.J., Maat M.J., Rep M., Grivell L.A.;
RT "The yeast nuclear gene MRF1 encodes a mitochondrial peptide chain
RT release factor and cures several mitochondrial RNA splicing
RT defects.";
RL Nucleic Acids Res. 20:6339-6346(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=S288C / F11769;
RX MEDLINE=97197983; PubMed=9046099;
RA Voet M., Defoor E., Verhasselt P., Riles L., Robben J., Volckaert G.;
RT "The sequence of a nearly unclonable 22.8 kb segment on the left arm
RT chromosome VII from Saccharomyces cerevisiae reveals ARO2, RPL9A,
RT Tipl, MRF1 genes and six new open reading frames.";
RL Yeast 13:177-182(1997).
CC -!- FUNCTION: MITOCHONDRIAL PEPTIDE CHAIN RELEASE FACTOR THAT DIRECTS
CC THE TERMINATION OF TRANSLATION IN RESPONSE TO THE PEPTIDE CHAIN
CC TERMINATION CODONS UAA AND UAG.
CC -!- SUBCELLULAR LOCATION: Mitochondrial.
CC -!- SIMILARITY: BELONGS TO THE PROKARYOTIC AND MITOCHONDRIAL RELEASE
CC FACTORS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X60381; CAA42932.1; -
DR EMBL: X99960; CAA68219.1; -
DR EMBL: Z72665; CAA96855.1; -
DR PIR: S28602; S28602.

```

```

DR SGD; S000311; MRF1.
DR InterPro: IPR000352; Pep_rel_factor_1.
DR Pfam: PF00472; RF-1; 1.
DR PROSITE: PS00745; RF_PROK_1; 1.
KW Protein biosynthesis; Mitochondrion; Transit peptide.
FT TRANSIT 1 413 MITOCHONDRION (POTENTIAL).
FT CHAIN 1 413 PEPTIDE CHAIN RELEASE FACTOR 1.
SQ SEQUENCE 413 AA; 46770 MW; AAMD4829748C7604 CRC64;

alignment_scores:
Quality: 95.00 Length: 307
Ratio: 0.642 Gaps: 14
Percent Similarity: 48.208 Percent Identity: 21.824

alignment_block:
US-09-839-894-9 x RF1M_YEAST
Align seg 1/1 to: RF1M_YEAST from: 1 to: 413

268 AGCAGTTCATCGTCACGCGTGAAACAAT.....ATACATTACAATT 311
:::|||||:::
30 ThrSerThrThrThrAsnSerLysSerAsnGlySerIleProThrGlnTyr 46
312 TACGGAAGAAAGAAAGTTTAATA.....AAAAGAGAGCTAC 346
:::|||||:::
46 rThcGluLeuSerProLeuLeuValLysGlnAlaGluLysrGluAlaG 63
347 AAATTAAGAGCTATTAACAATATTATGTTCAAAAGTGTAACTGCCATCC 396
:::|||||:::
63 IuLeuLysAsp.....LyuAspLysAspLeuSerCys..... 73
397 GCCTTAACACTTACATGAGCTCATTTTACATGATATAA..... 435
74 .....GlyIleHisPheAspValAsnLysGlnLysHisTyr 85
436 .....AACGCGGCTTCAGCTGCAAGTTTATTTATTTATTTCTCTG 475
85 rAlaLysLeuSerAlaLeuThrAspThrPheIleClnTyrLysGluLysL 102
476 CTGGCGCACTTAACAAATTTTGCCCTTTGGTGATCTGGATGCTACTCTG 525
:::|||||:::
102 euAsnGluLeuLysSerLeuGlnGluMetIleValSerAspProSerLeu 118
526 AAGTTAAGAGTAAAGAAAGCATATAGTAG.....ACCTTAGGAAC 566
:::|||||:::
119 ArgAlaGluAlaGluGlnGluTyrAlaGluLeuValProGlnTyrGluTh 135
567 TTACACTATTAATATACATTTAATTAACTGATTAAGGAATATTCAGA 616
:::|||||:::
135 rThSerSerArgLeuValAsnLysLeuLeuPro..... 146
617 TATGTTACCTCAGTTCAAAAGTGACGCTGCGTCATCTTAACCTGGT 666
:::|||||:::
147 .....ProHisProPheAlaAspLysProSerLeuLeuGluLeuArg 160
667 CCAACTGGTGGGGGC.....ACATATATGGAAGAAATCTGTGTA 707
:::|||||:::
161 ProGlyValGlyGlyIleGluAlaMetIlePheThrIleAsnLeuAs 177
708 TATGCTCTTTTATGATGATATAGTACTAACAGCAGCTTTTGGAGATAA 757
:::|||||:::
177 pMet.....TyrIleGlyTyrAla...AsnTyrArgLysTyrLysTyr 191
758 GATTTCAGATTAACAATCTTAATCTGATGGAAGAAATTTATTCAGAA 807
:::|||||:::
191 rGlyIleSerLysAsnGlnAsnGluSerGlySerGlyIleIleAspAla 207
808 ATTAAGATGACACCAAGAAATTCATATACCTTTCACCTCTGCTGCTG 857
:::|||||:::
208 IleuSerIleGluGlnAlaGlySerTyrAspArgLeuArgPheGlnAl 224
858 GGGTAAA.....AGTTAACTCCAAACAAATGAAAGCT 889

```



```

85 GATGAAGACATTAATATTTTGGCCCGGTCAGAGAAAGATCTTC 134
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
260 AspergillusleuaspGluLeuSerGlyValGluArgSerasp...Ly 275
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
135 CCCCAACATATATATTAATAC.....CATATTA 166
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
275 sValSerTyrLeuGluValAsnGluGlyTyrLeuLeuHisLeuT 292
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
167 CAGCA.....TACAGTAAAGTCATACCTGCTAT..... 195
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
292 hValAspGluAsnTyrLeuLysAspProGluValTyrProValSer 308
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
196 ...GATAGATGACTTTTATGTTGCTCTTCACATACATACCTAATG 242
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
309 IleAspSerThrSerLeuSerValSerSerAspThrPheValMetSe 325
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
243 AGCATGTCACACGATGAGATCCTACAGTTCATCGGTGCGGTGAA 292
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
325 ValTyrProThrThr...AsnTyrSerAlaSerSerGlnTyrStrpAsp 341
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
293 CAATATATACATTAATTTACGGAAGAAAGAGTAAATTAAGAGAG 342
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
341 laasn..... 342
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
343 CTACAAATTAAGGCTATAAACAATATTTGTTCAAAAGTTAACTGCC 392
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
343 ...LeuLysAlaTyr..... 346
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
393 ATCCGGCTAACACTTAACCTACGCTCATTTTAAGTAAATAAACCGG 442
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
347 .....ValLeuLysThrGlyTyr.....AspLysT 356
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
443 CTTCAGGTGACAGTTTATATTTATATTCCTGCGGCACTAAATAAT 492
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
356 hThrThrGlyThrAsn...TyrAlaPheMetLysPheAsnAsnLeuLysPro 371
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
493 TTGCGTTTGGTGGTATCTGGGATGCTACTCTGAAGTTAAGATAAAG 542
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
372 IleGlnAsnMetThrValThrLysAlaThrLeuLysThrTyrValAlaHis 388
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
543 AGCATATAGTAGACCTATGAGACTTACACTATAATATCTCTTAAT 592
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
388 sSerTyr.....TyrGlyThr.....LysA 395
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
593 TAACTGTAAGGAAATATTCAGATATGCTTACTCTCAAGTCAAAAGT... 639
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
395 laThrGly.....LeuTyrPheAspThrValAsnSerAsn 406
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
640 .....GACGCTGCGTCGATCTTAAGTGGCTCAACTGGTGGGCGAC 683
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
407 TyrAspAsnAlaLysValThrTyrPheAsnThrLysProAlaSerLysAsn.. 422
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
684 ATATATTGGAAGAAATCTGTGATATGCTTTATGATGATAT... 729
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
423 ...IleGlyLysAlaAspValHisLysGlyGlnTyrAlaSerTyrAspV 438
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
730 .....AGTACTAACAGACGCTCTTGGAGATTAAGATTT 762
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
438 alThrAlaAlaValLysSerTyrPheAsnSerGlyGlyAlaAsnTyrGlyPhe 454
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
763 CAGGATTAACATCTTAATCTGATGGAATTTATCTAAGAAATAA... 810
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
455 Lys...LeuHisThrAsnGlyAsnGlyLysGluTyrTrpLysLysLeuIl 470
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
811 .....AATGATGACACCAAGAAATTCATATCTT 841
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
470 eSerSerAlaAsnSerAlaAsnLysProTyrIleGluValThrTyrThrI 487
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
842 TGTCACCTTCTTGGCGGTTAAAGTTTAACTCAACA..... 879
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
487 lePro.....LysGlyAsnThrProThrIleLysAlaTyr 498
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
880 ...AATGGAACGTCATTAAATATTGCTGACGACAGCTTCTCTGAAACAA 926

```

```

seq_name: SwissProt_40:N100_YEAST
seq_documentation_block:
ID N100_YEAST STANDARD: PRT: 959 AA.
AC 002629;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Nucleoporin NUP100/NSP100 (Nuclear pore protein NUP100/NSP100).
GN NUP100 OR NSP100 OR YKL068W OR YKLJ36.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93054906; PubMed=1385442;
RA Wente S.R., Rout M.P., Blobel G.;
RT "A new family of yeast nuclear pore complex proteins.";
RL J. Cell Biol. 119:705-723(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=5288C;
RX MEDLINE=94378724; PubMed=8091863;
RA Rasmussen S.W.;
RT "Sequence of a 20.7 kb region of yeast chromosome XI includes the
RT NUP100 gene, an open reading frame (ORF) possibly representing a
RT nucleoside diphosphate kinase gene, tRNAs for His, Val and Trp in
RT addition to seven ORFs with weak or no significant similarity to
RT known proteins.";
RL Yeast 10:569-574(1994).
CC -!- FUNCTION: ESSENTIAL COMPONENT OF NUCLEAR PORE COMPLEX.
CC NUCLEOPORINS MAY BE INVOLVED IN BOTH BINDING AND TRANSLLOCATION OF
CC THE PROTEINS DURING NUCLEOCYTOPLASMIC TRANSPORT.
CC -!- SUBCELLULAR LOCATION: Nuclear pore complex.
CC -!- DOMAIN: CONTAINS G-L-F-G REPEATS.
CC -!- SIMILARITY: BELONGS TO THE GLFG FAMILY OF NUCLEOPORINS.
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to licenses@isb-sib.ch).
CC
DR EMBL: Z15035; CAA78753.1; -
DR EMBL: X75780; CAA53406.1; -
DR EMBL: Z28068; CAA81905.1; -
DR PIR: B44402; B44402.
DR PIR: S39173; S39173.
DR PIR: S44518; S44518.
DR SGD: S000151; NUP100.
DR InterPro: IPR004325; Nucleoporin_fg.
DR Pfam: PF03093; Nucleoporin_fg; 24.
KW Nuclear protein; Transport; Repeat.
FT DOMAIN 33 571
FT 29 X 6 AA APPROXIMATE REPEATS OF
FT G-L-F-G.
SQ SEQUENCE 959 AA; 9998 MW; D3985F9901BBA51 CRC64;

```

```

alignment_scores:
Quality: 94.50
Ratio: 0.480
Percent Similarity: 47.073
Length: 410
Gaps: 17
Percent Identity: 19.512

```

alignment_block:
US-09-839-894-9 x N100_YEAST ..
Align seg 1/1 to: N100_YEAST from: 1 to: 959

```

91 AGCATTAATTAATTTTGGCCGCGTGACAGAGCAATCTTCC..... 135
   ||| : : : : : ||| : : : : : ||| : : : : :
284 SerAsnSerGlyLeuPheGlyGlnAsnSerMetAsnSerSerThrGlnG1 300
136 .....CCCAACATTAATATTTA.....ATAACCATTA 163
   : : : : : ||| : : : : : ||| : : : : :
300 yValPheGlyGlnAsnAsnGlnMetGlnIleAsnGlyAsnAsnAsnA 317
164 TTACAGTACAGTGAAGTCACTCTGTATGATGATGATGATGATGATGAT 213
   : : : : : : : : : : : : : : : : : : : : :
317 snSerLeuPheGlyLysAlaAsnThrPheSerAsnSerAlaSerGly 333
214 TGTGTGCTTCTCACAATATACATTAATGAGACATGTCACACAGTGA 263
   : : : : : : : : : : : : : : : : : : : : :
334 LeuPheGlyGlnAsnAsnGlnGlnGlnGlnGlnGlnGlnGlnGlnG 350
264 TCCTAGCAGTTCATCGCTACGCT..... 288
   | : : : : | : : : : | : : : : | : : : : |
350 nSerGlnThrSerGlySerGlyLeuPheGlyGlnAsnAsnGlnLysG 367
289 .....GAAACAATATATACATTAATGATGATGATGATGATGATGAT 315
367 InProAsnThrPheThrGlnSerAsnThrGlyIleGlyLeu...PheGly 382
316 GAAAAAAGCACTTAATTAATAAAGAGCTACCAATTAAGCTTAATAA.. 363
   : : : : : : : : : : : : : : : : : : : : :
383 GlnAsnAsnAsnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnG 399
364 .....C 364
399 cAlaGlyThrThrGlySerLeuPheGlyGlyAsnSerSerThrGlnProA 416
365 AATATTGTGCAAAAGTCACTAAGCCGCGCTAAGCACTTAACATCA 414
   : : ||||| : : : : : ||| : : : : :
416 snSerLeuPheGlyThrThrAsnValProThrSer..... 427
415 GCTCATTTTAAGTCAATATAAAGCGGCTCAGTGCAAGTTTATATT 464
   ||| : : : : : ||| : : : : : ||| : : : : :
428 .....AsnThrGlnSerGlnGlnGlnGlnGlnGlnGlnGlnGlnG 438
465 ATATATTCCTGCTGCGCAATTAATAATTTGCTTTGCTGCTATCTGG 514
   ||| : : ||| ||| : : ||| ||| : : |||
439 .....GlyAlaThrLysLeuThrAsnMetProPheGlyGlyAsnProT 453
515 ATGCTACTGAAGTGAAGTAAAGACATATAGTGAACCTATATGA 564
   ||| : : : : : : : : : : : : : : :
453 hValA.....AsnGlnSerGlySerGlyAsnSerLeuPheGly 465
565 ACTTACACTATAATATCATATTAAATTAAGTGAAGAAATATATCA 614
   ||| : : : : : ||| : : : : : ||| : : : : :
466 ThrLysProAlaSerThrThrGlySerLeuPheGlyAsnAsnThrAla 482
615 GATATGTTACCTCACTCAAAAGT..... 639
   : : ||||| : : : : : : : : : : :
482 rThrThrValProSerThrAsnGlyLeuPheGlyAsnAsnAlaAsnAs 499
640 .....GACGCTCGCGTGCATCTTAC 660
   ||| : : ||| ||| : : ||| ||| : : |||
499 erThrSerThrThrAsnThrGlyLeuPheGlyAlaLysProAspSerGln 515
661 TTGGCTCAACT...GGTGGGCGACATAT..... 687
   : : ||||| : : ||||| : : : : :
516 SerLysProAlaLeuGlyGlyGlyLeuPheGlyAsnSerAsnSerAsn 532
688 .....ATTGGAAGAAAT.....TCGTGTGATA 709
   ||||| : : ||| ||| : : ||| ||| : : |||
532 rSerThrIleGlyGlnAsnLysProValPheGlyGlyThrThrGlnAsn 549

```

```

710 TGTGCTTTATGATGATATAGTACTTAACAGCAGCTTTG..... 750
   : : : : : : : : : : : : : : : : : : : : :
549 hrGlyLeuPheGlyAlaThrGlyThrAsnSerSerAlaValGlySerThr 565
751 ..GAGATTAAGATTTTCAAGATTAACAATCTTAATCTGATGGAAATTTA 797
   : : : : : ||| : : : : : ||| : : : : :
566 GlyLysLeuPheGlyGlnAsnAsnAsnThrLeuAsnValGlyThrGlnAs 582
798 TCTAAGGAATAATATGATGACACCAAGAAATTCATATACTTTGTGAC 847
   : : : : : ||| : : : : : ||| : : : : :
582 nValProValAsnAsnAsnThrThrGlnAsn..... 592
848 TTCCTGTGCGGCTTAAGTTTAAGTCAACAATATGATGATGATGATGAT 897
   ||| ||| : : : : : ||| : : : : : ||| : : : : :
593 ..AlaLeuLeuGlyThrThrAlaValPro.....SerLeuGln 604
898 ATTGCTGACGAGCTTCTGTGGAACAACATGGAATATGATGATGATGAT 947
   ||| : : : : : ||||| : : : : : : : : : :
605 GlnAlaProVal.....ThrAsnGlnGlnLeuPheSerLysI1 617
948 CACCATGCCA..GAATCAGTGTCCGCTGTGTGTGCTGACGCTT 994
   : : : : : ||| : : : : : ||||| : : : : :
617 eSerIleProAsnSerIleThrAsnProValLysAlaThrThrSerLysV 634
995 TGCAATGAGTGAACCAAGTGAAGAAATCCGAGCTGACCAATATATGAG 1044
   : : : : : ||| : : : : : ||| : : : : :
634 AlaAlaAlaAspMetLysArgAsnSerSerLeuThrSerAlaTyrArgLeu 650
1045 AATATTAAATGTTACTTTCACACCAAGTAGT 1074
   : : : : : ||| : : : : : ||||| : : : : :
651 AlaProLysProLeuPheAlaProSerSer 660

```

seq_name: SwissProt_40:LEU2_BUCRP

```

seq_documentation_block:
ID LEU2_BUCRP STANDARD; PRT; 471 AA.
AC P48573;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE 3-isopropylmalate dehydratase large subunit (EC 4.2.1.33)
DE (3-isopropylmalate isomerase) (Alpha-IPM isomerase) (IPMI).
GN LEUC.
OS Buchnera aphidicola (subsp. Rhopalosiphum padi).
OG Plasmid PRPE.
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX NCBI_TaxID=98793;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95333198; PubMed=7608990;
RA Bracho A.M., Martinez-Torres D., Moya A., Latorre A.;
RT "Discovery and molecular characterization of a plasmid localized in
RT Buchnera sp. bacterial endosymbiont of the aphid Rhopalosiphum
RT padi."
RL J. Mol. Evol. 41:67-73(1995).
CC -1- CATALYTIC ACTIVITY: 3-ISOPROPYLMALATE = 2-ISOPROPYLMALATE +
CC H(2)O (ALSO CATALYSES 2-ISOPROPYLMALATE + H(2)O = 3-HYDROXY-
CC 4-METHYL-3-CARBOXYPENTANONE).
CC -1- PATHWAY: SECOND STEP IN LEUCINE BIOSYNTHESIS.
CC -1- SUBUNIT: CONSISTS OF TWO DIFFERENT SUBUNITS: LEUC AND LEUD.
CC -1- SIMILARITY: BELONGS TO THE ACONITASE/IPM ISOMERASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X71612; CA50617.1; -
CC InterPro: IPR001030; Aconitase.
CC Pfam: PF00330; aconitase; 1.

```


FT CARBOHYD 98 98 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 164 164 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 252 252 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 412 AA; 45948 MW; E420D3530D1A04D9 CRC64;

alignment_scores:

Quality: 93.50 Length: 305
 Ratio: 0.645 Gaps: 14
 Percent Similarity: 47.341 Percent Identity: 18.361

alignment_block:

US-09-839-894-9 x THBG_SHEEP

Align seg 1/1 to: THBG_SHEEP from: 1 to: 412

```

97 ACTAATATTTTGGCCGCGTGACAGACAAATCTCCCAACATAA 146
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
155 ThrGluValPheSerThrAspPheSerAsnValSerAlaIleGlnGlu 171
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
147 TATATTAATACCATATTACAGCATACGTAAGT..... 183
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
171 uIle...AsnSerHisValGluArgGlnThrLysGlyIleValGlyL 187
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
184 .....CATACTCTGATGATAGATGACTTTTGA 213
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
187 euIleGlnAspLeuLysProAsnThrIleThrValLeuValAsnTrpLeu 203
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
214 TGTTCGCTTCACATATACATTAATGAGACATGCCACACAGTAGAA 263
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
204 CysPheLysAla...GlnTrpAlaAsnProPheAspProSerLysThrGlu 219
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
264 TCCTAGCAGTTCATCGCGTACGCGGTGAACAATAATTAACATTACAATT 313
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
219 uGluGlySerSerPheLeuValAspLysThrThr...ThrValGln.... 233
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
314 CGGAAAAAGAGTTTAAATAAAGAGACGACATAAATTAAGGCTATAAA 363
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
234 .....ValProMetMetHisGlnValAlaAspGlnTrpTyr 244
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
364 CAATTATTTGTTCAAAAGTGTAACTGCCATCCGCGCTAACACTTAAGTC 413
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
245 HisLeuValAspThrGluLeuAsnGly.....ThrVal 255
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
414 AGCTCATTTTAACGTGTAATAAAGCGCGCTTCAGGTCAAGTTTATAT 463
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
255 IleuGlnMetAspTyrSerLysAsnAla.....LeuAlaLeuPheV 269
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
464 TATATATTCCTGCTGCGCACTAAATAATTTGCCCTTTGGTGTATCTGG 513
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
269 aIleuProLysGluGlyGlnMetGluGlyVal..... 279
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
514 GATGCTACTCTGAAGTAAAGATAAAGACGATATAGTAGACCTATAG 563
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
280 GluGlyAlaMetSerSerLysIleLeuLysLysTrpAsn..... 292
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
564 AACTTACACTATAATATCACTATTAAATTAAGTAAAGGAAATATTC 613
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
293 .....ArgLeuLeuGlnLysGlyTrpValA 301
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
614 AGATATGTTACCTCACTCAAAAGTCAACGCTCGCGCTGACCTTAAGTC 663
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
301 snLeuPheValProLysPheSerIleSerAlaThrTyrAspLeu..... 315
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
664 CGTCCAACTGCTGG..... 678
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
316 .....GlyGlyIleLeuLeuLysMetGlyIleGlnAspAlaPheAl 329
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
679 .....GGCACATATATTTG 691
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
329 aAspAsnAlaAspPheSerGlyLeuThrLysAspAsnGlyLeuLysValS 346
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
692 GAAGAAATTCCTGTGATATGCTTTTATGATGATATAGTACTAAGAC 741
  
```

```

346 erAsnValAlaHisLysAlaMetPheTyrIleGlyGluLysGlyThrGlu 362
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
742 AGCTCTTGGAGATAGATTTTCAGATTAACAATCTTAATCT..... 783
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
363 AlaIleProGluValArgPheLeu...AsnGlnProGluThrThrLeuLe 378
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
784 .....GATGGAAATTTATCTAAGCAAAATTAATG 814
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
378 uHisProIleIleGlnPheAspArgSerPheLeuLeuIleLeuGluL 395
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
815 ATGACACCAAGAAATGCGATATACCTTTGTCACCTTCTCTGGCGGTAA 864
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
395 yAsnThrArg.....SerIleLeuPheLeuGlyLys 405
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
865 AGTTTAAGTCCACA 879
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
406 ValValAspProThr 410
  
```

This Page Blank (uspto)

401	TAACCTTAACTACAGCTCATTTTAACTGTAATTAACAAACGGCGCTTCAGG	450
134	euthrleuansnserrAlaHisphasnCysasnIysSnlaIaAsergIy	150
451	GCAAGTTTATATTATATATATCTCGCTGGCGAACAATAAAATTTGCCCTT	500
151	AlaSerleuTyrlleuTyrlleProIaIagIyGluIleuIysasnleuProh	167
501	TGGTGGTAACTGGGAGCGTAACTCTGAAAGTTAAGAGTAAAAAGCGATATA	550
167	egIyGlyIleTrrPaspAlaThrIleuIysleuAgyaIyLysArgAlTyAs	184
551	GTGAGACCTATGGAGAACTTACACTATATAATCATCTTAAATTAATCGAT	600
184	ercIuThrTyrgIyThTyrrThIleasnIleThrlIeIySleuThrrAsp	200
601	AAGGAATATTTACAGATATAGTTTACCTCAAGTTCAAAAGTACGCTCGCT	650
201	LysGIyAsnIleGlnIleTrrleuProGlnPhelYsserAspAlaArgIyA	217
651	CGATCTTAACCTGCGTCCAACTGGTGGGGGCACATATATTGGAAGAATT	700
217	LaspleuAsnleuArpProThrIyGlyGlyThTyrrIleGlyArgAsns	234
701	CTGTGTATATGTCTTTTATATGATGATATAGTACTAACAGACCTCTTGG	750
234	erAlaIaspMetCysPheTyrrAspGlyTyrrSerThrrAsnSerSerleu	250
751	GAGATTAAGATTTCAGATAACAATCCTAAATCGATGGGAAATTTTATCT	800
251	GluIleArpPheGlnAspAsnAsnProIySserAspGIyLysPheTyrlle	267
801	AAGGAATAATTAAGAGACACCAAAATTTGGATATCTGTTGCTGCCTTC	850
267	uArgIyIleAsnAspAspThrIySleuIleAlaTyrrIhrleuSerleu	284
851	TCITGGCGGGTAAAGTTTAACTCCACAACAATGGAACGTATTAAATATT	900
284	euleuAlaGlyLysSerleuThrProThrAsnGlyThrSerleuAsnIle	300
901	GCAGAGGCACTTCTCTCGGAACAACCTGGAATAGATTACAGCTTCAC	950
301	AlaAspAlaIaIaSerleuGlnThrrAsnTrrPaspArgIleThrAlaValh	317
951	CATGCCAGAAATACAGTTCGCGGTGGTGGTGGCCCTGGACGTTTGCAT	1000
317	TrMetProGlnIleSerValProValIeucYstrProGlyArgIeGlnh	334
1001	TGGATGCCAAAAGTGAATAATCCCGAGGCTGGACAATATATGGGTAAATTT	1051
334	euaSpAlaIySValGluAsnProGluAlaGlyGlnTyrrMetGlyAsnIle	350
1051	AATGTACTTTCACACCAGTAGTCAAAACATC	1083
351	AsnValThThrPheThrProSerSerGlnThrrIleu	361

```
seq_name: sp_bacteria:Q47115
seq_documentation_block:
ID Q47115 PRELIMINARY; PRT; 363 AA.
AC Q47115;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE COOD PRECURSOR.
GN COOD.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBL_TaxID=562;
RN [1]
RP
RZ
SEQUENCE FROM N.A.
MEDLINE=94344028; PubMed=7915003;
XX
```

RA Firehitch B., Karakashian A., Melsen L.R., Wakefield J.C., Scott J.R.
 RT "Cooc and Cooc are required for assembly of CSI pIII.",
 RL Mol. Microbiol. 12:387-401(1994).
 DR EMBL, X76908, CAA54230.1,
 DR InterPro: IPR001412; tRNA-synt.1,
 DR PROSITE; PS00178; AA_TRNA_LIGASE_I, UNKNOWN_1.
 KW Signal.
 RV Signal.
 FT CHAIN 1 18 1 POTENTIAL.
 FT CHAIN 19 363 COOD.
 SO SEQUENCE 363 AA; 40139 MM; AA9488AANDDED572 CRC64;

alignment_scores:	
Quality:	949.50
Ratio:	3.230
Percent Similarity:	79.245
	Length: 371
	Gaps: 11
Percent Identity:	53.100

alignment_block:
US-09-839-894-9 x Q47.115

Align seg 1/1 to: Q47115 From: 1 to: 363

[illegible]

```

621 |G|A|C|T|G|A|G|T|T|C|A|A|A|G|T|A|C|G|C|G|C|G|C|A|T|C|T|A|C|T|G|C|G|C|T|C|A|A| 670
210 |P|L|E|U|R|O|C|I|P|H|E|I|S|S|E|R|S|P|R|O|A|R|G|Y|A|L|S|P|L|E|U|A|N|S|L|E|U|A|G|P|R|O|I| 227
671 |C|T|G|T|G|G|G|G|C|A|T|A|T|A|T|T|G|A|A|A|A|T|T|C|T|G|T|G|A|T|A|T|G|C|T|T|T|A|T| 720
227 |L|E|G|I|A|S|N|T|Y|R|L|Y|S|T|Y|S|E|R|G|I|S|E|R|A|N|S|E|R|L|E|U|A|S|P|E|C|T|Y|S|H|E|T|Y|R| 243
721 |G|A|T|G|A|T|A|T|A|G|Y|A|C|T|A|A|C|A|C|A|C|C|T|T|T|G|G|A|G|A|T|A|G|A|T|T|C|A|G|A|T|A| 770
244 |A|S|P|G|I|Y|R|S|E|R|T|H|R|A|N|S|E|R|A|S|P|S|E|R|M|E|T|V|A|I|L|L|E|Y|S|P|H|E|G|I|A|S|P|A|S| 260
771 |C|A|N|O|C|T|A|A|A|T|C|G|A|T|G|G|A|A|T|T|T|A|T|C|T|A|A|G|A|A|A|A|A|A|A|T|A|T|G|A|T|G|A|C|A| 820
260 |P|A|N|P|H|O|T|H|r|A|N|S|E|R|G|I|U|T|R|A|N|S|E|U|R|Y|R|L|Y|S|I|L|E|G|Y|...G|L|Y|T| 276
821 |C|C|A|A|G|A|A|A|T|T|G|C|A|T|A|T|A|C|T|T|T|G|T|C|A|C|T|C|T|G|G|C|G|G|T|A|A|A|G|T|T|A| 870
276 |H|R|G|I|U|L|Y|S|L|E|U|P|R|O|T|Y|A|L|A|V|A|L|S|E|R|L|E|U|E|U|N|C|I|Y|L|U|L|Y|S|I|L|E|P|H|E| 292
871 |A|C|T|C|A|C|A|A|A|T|G|A|C|G|C|A|T|T|A|A|A|T|A|T|T|G|C|T|G|A|C|G|C|A|G|C|T|T|C|T|G|C|A| 920
293 |T|Y|R|P|R|O|V|A|L|A|N|G|I|N|S|E|R|P|H|E|R|H|I|L|E|A|S|A|S|P|S|E|R|S|E|R|V|A|L|L|E|G|I| 309
921 |A|A|C|A|A|C|G|G|A|T|A|G|A|T|T|A|C|G|T|G|C|A|C|G|C|A|C|C|A|G|C|A|A|A|A|C|A|G|T|T|C| 970
309 |U|T|R|A|N|T|P|A|N|R|Y|A|I|T|R|A|L|A|V|A|L|A|A|A|E|T|P|R|O|G|I|U|V|A|I|A|N|S|V|A|L|P| 326
971 |C|G|T|G|T|G|T|G|G|C|C|T|G|C|G|A|C|G|T|T|G|C|A|T|G|G|A|T|G|G|A|C|A|A|A|G|T|G|A|A|A|T| 1020
326 |R|O|V|A|L|L|E|U|C|Y|S|T|R|P|R|O|A|L|A|R|G|L|E|U|L|E|U|E|U|N|C|I|A|S|P|A|L|A|S|N|A| 342
1021 |C|C|G|G|A|G|C|G|T|G|A|C|A|A|T|P|A|N|G|G|A|A|T|A|T|A|N|G|T|A|C|T|T|G|A|C|C|A|G| 1070
343 |P|R|O|S|P|A|R|A|A|G|I|G|I|N|T|Y|R|S|E|R|G|I|N|I|L|I|E|R|Y|R|I|E|T|R|P|H|E|R|H|P|R|O|S|E| 359
1071 |T|A|G|T|C|A|A|C|A|C|T|C| 1083
359 |R|V|A|I|G|I|U|A|N|S|L|E|U| 363
seq_name: sp_bacteria:0471119
seq_documentation_block:
ID 047119 PRELIMINARY; PRT: 364 AA.
AC 047119;
DT 01-NOV-1996 (TREMBLERL 01, Created)
DT 01-NOV-1996 (TREMBLERL 01, Last sequence update)
DT 01-NOV-1998 (TREMBLERL 08, Last annotation update)
DE COTD PRECURSOR.
GN COTD.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
NCBI_taxid=562;
RX [1]
RC STRAIN=96071908; PubMed=7591145;
RA Froehlich B.J., Karakashian A., Sakelaris H., Scott J.R.;
RT "Genes for CS2 pilli of enterotoxigenic Escherichia coli and their
RT interchangeability with those for CSI pilli.";
RL Infect. Immun. 63:4849-4856(1995).
DR EMBL; Z47800; CAA87763.1; -.
KW Signal.
FT SIGNAL 1 18 POTENTIAL.
FT CHAIN 19 364 COTD.
SQ SEQUENCE 364 AA: 40979 MW: ECTAC738A0E195B6 CRC64;
```

```
alignment_scores:      Length: 366
      Quality: 930.50
      Ratio: 3.209
      Gaps: 3
```

```
Percent Similarity: 79.235 Percent Identity: 48.361
Alignment_block:
US-09-839-894-9 x Q47119 ..
Align seg 1/1 to: Q47119 from: 1 to: 364
```

```
alignment_block:
UC-09-8329-8004-0 v 047119
```

US-09-839-894-9 x Q47119

Align seg 1/1 to: Q47119 from: 1 to: 364

```

cent Similarity: 79.235      Percent Identity: 48.361
mmcnt_LbLock:
05-839-894-9 x 047119      ..
ign seg 1/1 to: 047119 from: 1 to: 364

1  ATGAAATGAAGATTATATATATTTTACATATGTTTCTCTAGACACTTT 50
1  ::::::::::::::::::::::::::::::::::::::::::::::::::::
1  MetcylsVal1llebheValleuSerMetPheLeuCySsSerGlnVal17
51  TACATTTGGCGTATGGCGAGATAAATTCGCCGAGATGAAGACATTA 100
17  rGlyGlnSerTThrPHisThrAsnValGluLacIlySerIleasnlySthr 34
17  ::::::::::::::::::::::::::::::::::::::::::::::::::::
34  lserIleGlyProIleAspArgSerAlaIleAlaIleSerTyrProAlHis 50
145  AATATATTAATATACATATTTTACAGATACAGTAAAGTACTACTGTGA 194
145  ::::::::::::::::::::::::::::::::::::::::::::::::::::::
51  TyxIlePheHisGlnHisValaGlyTyrAsnlySAspHisSerLeuPh 67
195  TGAATAGATACACTTTTATGTGTGGTCTTCACAAATACACTTAATGGAG 244
67  eAspArgMetThrPheLeuCySMeSerSerThrAspAlaSerlyScIyA 84
245  CAGTCCCAACAGATGAGAAATCTTACACTGATCGATGACGGGGAACA 294
84  lacyProthrGlyAsnSerlySserSer.....GlnGlyGluThr 98
295  AATATACATTAACATTTATACGAAAAAGAAAGATTATATAAAGAGAGCT 344
99  AsnIleLySerLeuPhePheThrGluLySsSerLeuAlaArglySthrLe 115
345  ACAATTAAGGCTAATAACATTTATGTGTGCAAAAGCTTAACGCG... 390
115  uasnleLySgLyTyrLyAspPheLeuLyGlnIserAspArgLySleH 132
391  ....CCATCCGGCCCTAACACTTAACAGCTACATTTTAACTGAATAAA 435
132  lstrYAlaAspLySMeLAsnLeuAsnSerHisThrValLyCySValGly 148
436  AACCGCGCTCAGGTGCAAGATTATATATTATATATCTCTGTCGCAACT 485
149  SerPheThrArgIyAlaSerPheThrLeuTyrlleProGlnGlyGluI 165
486  AAAAAATTTGCCCTTTGGTGGTATCTGGAGTCTACTGTGAAGTAAAGAG 535
165  eAspGlyLeuLeuThrGlyLyIleTrpGluAlaThrIleuGlnIleuArgV 182
536  TAAAAAGACATTAAGTACAGCTTGAAGAACTTACATATATATATAGACT 585
182  allysArgHisTyAspTyArgAsnHisGlyThrTyLySValAsnIleThr 198
586  ATTAAATTAAGTATGAAGGAAATATTCAGATATGGTATACCTCAGTTCAA 635
199  ValAspLeuThrAspLySgLyAsnIleGlnValIleThrProllysPheH 215
636  AAGTGAAGCTCGCGTGCATCTTAACCTGGCGTCAACTGTGTGGGCGACAT 685
215  sSerAspProArgIleAspLeuAsnLeuArgProGluIleLySsnIlyust 232
686  ATATTAAGAAATTCGTGTGAATATGGTCTTTATGATGAGATATAGTACT 735
232  ySerGlySerAsnValaLleuGlnMetCySLeuTyAspGlyTyrSerThr 248
736  AACAGCAGACTCTTTGGAGATAAGATTATACAGATTAACAATCTTAATCTGA 785
249  HisSerGlnSerIleGluMetArgPheGlnAspAspSerGlnThrGlyAs 265
786  TGGCAATTTTATCTAGAGAAATTAATGATGACACCAAGAAATTCGAT 835

```

265 nansngltyrAsnleuIlelysthrGlyuproLeuLysLysleuProt 282
836 ATACTTGTACCTCTCTGGGGGTTAAAGTTAACTCCAGCAAACTGA 885
282 yrlYsleuSerleuLeuGlyLargGluPheTyrProAsnGly 298
886 ACCTCATTAATATGCTGACGAGCTCTCTGAAACAACTGAATAG 935
299 GluAlaPheThrIleAsnAspThrSerSerleuPheIleAsnTrpAsn 315
936 AATTACAGCTGTACCATGCGAGAAATCAGTTCGGGTGTGTGGC 985
315 gIleYsSerValSerleuProGlnIleSerIleProValleuCysTrp 332
986 CTGAGCTTGTGATGATGATGAAAGTGAATCCGAGCTGGACAA 1035
332 roAlAsnleuThrPheMetSerGluLeuAsnProGluAlaGlyGlu 348
1036 TATATGGGTAAATATTAATGTTACTTTCACACCAAGTACCAACTC 1083
349 TyrSerGlyIleLeuAsnValThrPheThrProSerSerSerleu 364

seq_name: sp_bacteria:Q9XDS1

seq_documentation_block:

ID Q9XDS1 PRELIMINARY; PRT; 359 AA.

AC Q9XDS1:
DT 01-NOV-1999 (TREMblrel. 12. Created)
DT 01-NOV-1999 (TREMblrel. 12. Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17. Last annotation update)
DE TSAD PROTEIN.
GN TSAD OR TCFD.
OS Salmonella typhi, and
OS Salmonella enterica.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmoneila.
OX NCBI_TaxID=601, 28901;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S. typhi; STRAIN=GFU10007;
RA Hashimoto Y.;
RT "Identification of a putative fimbrial operon of Salmonella typhi";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=S. enterica; STRAIN=RKS 3333;
RX MEDLINE=99348391; PubMed=10417651;
RA Foljesson A., Advani A., Sukupoly S., Pfeifer J.D., Normark S.,
RA Lotfah S.;
RT "Multiple insertions of fimbrial operons correlate with the evolution
RT of Salmonella serovars responsible for human disease.";
RL Mol. Microbiol. 33:612-622(1999).
DR EMBL: AB029403; BAA82272.1;
DR EMBL: AJ242964; CAB51577.1;
SQ SEQUENCE 359 AA; 39741 MW; BB88AAE9A62052CE CRC64;

alignment_scores:

Quality: 275.50 Length: 303
Ratio: 1.514 Gaps: 11
Percent Similarity: 60.066 Percent Identity: 29.703

alignment_block:

US-09-839-894-9 x Q9XDS1 ..

Align seg 1/1 to: Q9XDS1 from: 1 to: 359

208 TTTTATGTTGTTCTCTCACAATACCTTAATGAGCATGTCCAAACAG 257
:::||||| |||:::|||||
79 TTPValCysArgSerAsnArgAsnGluAsnGlyAlaCysGluGlu 95
:::||||| |||:::|||||
258 TGAGAAATCTAGCATTCATGGTGGTGAACCAATATATCAATTAC 307
:::||||| |||:::|||||

95 rHsleuValTrpTrpTyrAlaPheGluValaTyrSerLysIleArgLeuA 112
308 AATTACGGAAGAAAGAGTTTAAATAAGAGAGCTACCAATTAAAGC 357
112 rPheArgGluGlnIleSerHisAlaGluIleThrLeu..... 124
358 TATAAACAATTAATTTGCAAAAGTGTAAAC.....TGCCATCCGGCCT 401
125IleLeuLeuGlySerValArgAspAlaCysTyrThrCysLys 138
402 AACACTTAACAGCTCATTTTAACTGTAAATAAAGCGGCTCA.... 447
138 I.....IleAsnMetAsnAlaAlaAlaCysG 147
448GGTGAAGTTATATTAATATATCTGCTGCTGCGCACTAAAAAT 492
147 IIntProIArgSerleuLysleuArgIleProSerGluGluLeuAlaLys 163
493 TTGCTTTTGTGGTATCTGAGATGCTACTGTGAATTA...AGACTAA 539
164 IleProThrSerCylThrTrpLysAlaThrIleuValIleuAspTyrLeuI 180
540 AAGACGATATAGTAGAGACTATGGAACCTTACACTATAATATACATATA 589
180 nTrpGluGluAspAspProleuGlyThrSerThrThrAspIleThrLeuA 197
590 AATTACTGAT.....AAGGAAATATCAGATATAGTATGCTGCTAGTTC 633
197 snValThrAspHisPheAlaGluAsnAlaAlaIleTyrThrProGlnPhe 213
634 ...AAAAGTACGCTGCTGCTGCTGCTTAACTTGCCTGCACTGAGGGG 680
214 GluThrAlaThrProAlaGluAsnleuIleArgMetAsnAlaLe 230
681 CACATATATGGAAGAAATTTCTGATATGCTGCTTATGATGATATA 730
230 rGlnMetSerGluArgAlaAsnleuAspMetCysLeuTyrAspGlu...G 246
731 GTACTAAGACGAGCTCTTGAAGATGAATTCAGATGAACAATCTTAA 780
246 lYValLysAlaArgSerleuGlnMetLysIleGluGlySerAsnLysSer 262
781 TCTGATGGGAATTTTATCTAAAGAAATTAATGATGACACCAAGAAAT 830
263 GluThrGlyPheGlnValIleuYsSerAspSerAlaAspThr..... 277
831 TGCATATACTTTGTCACTCTCTGCGGGGTAAGATTTAACTCCACAA 880
277 eAspTyrAlaValSerMetAsnTyrGlyArgSerIleProValThrA 294
881 ATGGAACGTCATTAATATTTGCGAGCGAGCTTCGTGAAACAAACTCG 930
294 rGluValAlaGluPheSerleuAspAsnValAspIysAlaIleThrArg... 309
931 AATGAAATTAAGCTGTCAACAGCAGCAAGAAATCAGTTCGCTGTGTG 980
310ProValValleuProGlnArgGlnAlaValArgCys 322
981 TTGGCCTGGACGTTTGCAATGTGATGCAAA.....GTGGAATTC 1021
322 sValProValProleuThrThrThrGlnProPheAsnIleArgGluL 339
1022 CCGAGGCTGGACAATATATGGGTAATATTAATGTTACTTTCACACCAACT 1071
339 ysArgSerGluGluTyrGlnGlyThrLeuThrValThrMetLeuMetGly 355

1072 ACTCAACA 1080

356 ThrGlnThr 358

seq_name: sp_bacteria:P71132

seq_documentation_block:

ID P71132 PRELIMINARY; PRT; 847 AA.
 AC P71132;
 DT 01-FEB-1997 (TREMblrel. 02, Created)
 DT 01-FEB-1997 (TREMblrel. 02, Last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE POMp1A.
 OS Chlamydia abortus.
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxID=83555;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=526/3;
 RX MEDLINE=96406378; PubMed=8810511;
 RA Longbottom D., Russell M., Jones G.E., Lainson F.A., Herring A.J.;
 RT "Identification of a multigene family coding for the 90 kDa proteins
 of the ovine abortion subtype of Chlamydia psittaci.";
 RL FEMS Microbiol. Lett. 142:277-281(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=526/3;
 RX MEDLINE=98187897; PubMed=9529048;
 RA Longbottom D., Russell M., Dundar S.M., Jones G.E., Herring A.J.;
 RT "Molecular cloning and characterization of the genes coding for the
 highly immunogenic cluster of 90-kilodalton envelope proteins from the
 Chlamydia psittaci subtype that causes abortion in sheep.";
 RL Infect. Immun. 66:1317-1324(1998).
 DR EMBL: U65942; AAC15921.1;
 DR InterPro: IPR003368; DUF145.
 DR InterPro: IPR003357; OMP.
 DR Pfam: PF02415; DUF145; 1.
 DR Pfam: PF02385; OMP; 1.
 SQ SEQUENCE 847 AA; 90694 MW; 754C958E7F11179E CRC64;

alignment_scores:
 Quality: 117.50 Length: 441
 Ratio: 0.573 Gaps: 22
 Percent Similarity: 46.485 Percent Identity: 21.315

alignment_block:
 US-09-839-894-9 x P71132 ..

Align seg 1/1 to: P71132 from: 1 to: 847

```

15 ATTTATTTTACATGTTTCTCTCAGTACTTTTACATTCGCTGAT 64
   :::::::::::::::::::: |||:::
5 ValTyrTrpPheLeuIle.SerSerSerLeuIleValSerSerSerLeuT 21
   :::::::::::::::::::: |||:::
65 CGGAGATAAATCCCGGAGATGAAGCATTAATATTTTGGCCCG 114
   :::::::::::::::::::: |||:::
21 YrSerGluGlu...ProAspIleYrThrLeuThrSerAlaHisSerTyr 36
   :::::::::::::::::::: |||:::
115 CGTACAGCAAGCATCTTCCCAACAATATATTAATAACATAT 164
   :::::::::::::::::::: |||:::
37 AsnGlyAsnThrAsnSerGluPro.....PheAsnProLe 48
   :::::::::::::::::::: |||:::
165 TACAGCATACAGTGAAGTCACTCTGTATAGATGAGTCTTTTAT 214
   :::::::::::::::::::: |||:::
48 userThrSerAsnSerAsnGlyThrIleTyrThrCysThrGlyAsnIleC 65
   :::::::::::::::::::: |||:::
215 GTTGTGCT.....TCTCAATACACTTAATGAGCATGCCA 252
   :::::::::::::::::::: |||:::
65 YrIleAlaTyrAlaGlyLeuAspGlySerGlyLeuSerSerSerCysPhe 81
   :::::::::::::::::::: |||:::
253 ACC..... 255
   |||
82 ThrAspThrAlaGlyAsnLeuSerPheLeuGlyAsnGlyTyrThrLeu 98
   |||
256 .....AGTGAATCCTTACGACATTCATCGGTCA 283
   |||
98 sPheAspAsnIleThrThrGlnSerSerSerHisProGlyAlaIleSerVal 115
   |||
284 GCGGTGAACAATAATATACATTTTACGGAAGAAAGATTATATA 333

```

```

115 ergLy...ThrAsnLysThr..... 120
   ||||| ||||| |||
334 AAAAGAGAGCTACAAATTAAGGCTATAAACATATTTGTTCAAAAGTGT 383
   |||||:::|||||:::
121 .....LeuAspIleSerGlyPheSer.....LeuPheSerCysAl 132
   :::::::::::::::::::: |||:::
384 TAACTGCCCATCCCGCCCTAACACTTAACCTACGT..... 417
   ||||| ||| |||
132 aTyrCysProProGlyAlaThrGlyTyrGlyAlaIleLysAlaValGly 149
   :::::::::::::::::::: |||:::
418 .....CATTTTAACTGTAAT 432
   ||| |||||:::
149 snThrThrIleLysAspAsnSerSerLeuValPheHisLysAsnCysSer 165
   :::::::::::::::::::: |||:::
433 AAAAAGCGGCTTCAGGTGCAAGTTTATATTATATATATCTGCTGGCGA 482
   :::::::::::::::::::: |||:::
166 ThrGlyGluGlyAlaIleGlnCysLysAlaSerSerSerGluAlaGly 182
   :::::::::::::::::::: |||:::
483 ACTAAA.....AATTGGCCCTTT..... 501
   ||||| ||| |||
182 uLeuLysIleGluAsnAsnGlnAsnLeuValPheAlaGluAsnSerSers 199
   :::::::::::::::::::: |||:::
502 .....GGTGTATCTGGGATGCTACTCTGAAGTTAAGATAAAGA 543
   ||||| ||| |||
199 erSerSerGlyGlyAlaIleTyrAlaAsp...LysLeuThrIleValSer 214
   :::::::::::::::::::: |||:::
544 CGATATAGTAGACCTATGACATTAACATTAATAATATCACT..... 585
   :::::::::::::::::::: |||:::
215 GlyIlyProThrLeuPheSerAsnSerValSerAlaSerProly 231
   :::::::::::::::::::: |||:::
586 .....ATTAATTAATGATTAAGGAAATATTCAGATATGCTTAC 625
   |||::: |||::: |||:::
231 sGlyGlyAlaIleCysIleLysAspSerGlyGly...GluCysSerLeuT 247
   :::::::::::::::::::: |||:::
626 CTCAGTTCAAAAGTAGCGCTCGCTCGATCTTAAC.....TTCGCTCA 669
   :::::::::::::::::::: |||:::
247 hrAlaAspLeuGlyAspIleThrPheAspGlyAsnLysIleIleLysThr 263
   :::::::::::::::::::: |||:::
670 ACTGGTGGGGCACATATATTTGAAGAAATTCGTGTATATG..... 711
   :::::::::::::::::::: |||:::
264 AsnGlyGlySerProThrValThrArgAsnSerIleAspLeuGlySerSe 280
   :::::::::::::::::::: |||:::
712 .....TGCTTTT 718
   |||
280 rGlyLysPheThrLysLeuAsnAlaLysGluGlyPheGlyThrPhePheT 297
   |||
719 ATGATGGATATAGT...ACTAACAGACGCTCTTGGAGATTAAGATTTCAG 765
   |||||::: |||::: |||:::
297 YrAspProIleThrGlyGlyGlySerAspGluLeuAsnIleAsnLysGln 313
   :::::::::::::::::::: |||:::
766 GATAACAATCCTTAATCTGATGGGAAA.....TTTATATCAAGAAAT 809
   |||||::: |||::: |||:::
314 AspThrValAspTyrThr...GlyLysIleValPheSerGlyGluArgLe 329
   :::::::::::::::::::: |||:::
810 AAATGATGACACCAAGAAATTCATATCTTG..... 843
   :::::::::::::::::::: |||:::
329 userAspGluGluLysValAlaAlaAsnLeuLysSerAspPheLysG 346
   :::::::::::::::::::: |||:::
844 .....TCACTTCTCTTGGCGGT..... 861
   ||||| |||
346 InrProLeuLysIleGlySerGlySerLeuIleLeuLysAspGlyValThr 362
   |||
862 .....AAAGTTTAATCTCAACAATGAAGCATTAATAATATTCG 902
   |||||::: |||::: |||:::
363 LeuGluThrLysSerPheThrGlnThrGluGlyAlaThrValAlaMetAs 379
   :::::::::::::::::::: |||:::
903 TGACGCACTTCTCTGGAACA.....ACTGGAATATGAATTTACAGCTG 946
   :::::::::::::::::::: |||:::
379 rLeuGlyThrThrLeuGlnThrProSerSerGlyGlyLysThrIleThrL 396
   |||
947 TCACCATGCCGAATATGATGTTCCGGTGTGTGCTGACCTGACGTTTG 996
   |||

```



```

342 ysserlyrhrhysglnproleuysllleglyalaglyserleuvalleu 358
844 .....TCACCTCTCTGGGGGTTAAAGATTAACTCCAAATAAGAAC 887
      :::: ||| :|||:::||||| |||::|::|
359 LysAspGlyValThrLeuGlnLysLysLysLysLysLysLysLysLys 375
      :|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
888 GTATTATTAATTTGCTGACGACGAGCTTCTCTGGAACAACAATCGAT 937
      :::: ||| :|||:::||||| |||::|::|
375 rTrnValValMetAspLeuGlyThrThrLeuGlnThrProSerSerG 392
      :|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
938 TTACAGCTGTCACCAAGCCAGAAATCACTGTTCCGGTG 975
      :|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
392 LysIleThrIleThrIleThrAsnLeuAspIleAsnIle 404

```

seq_name: sp_bacteria:P94750

```

seq_documentation_block:
ID      P94750      PRELIMINARY;          PRT;      2349 AA.
AC
DT      01-MAY-1997 (TREMBLrel. 03, Created)
DT      01-MAY-1997 (TREMBLrel. 03, Last sequence update)
DT      01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE      INVASIN (FRAGMENT).
OS      Escherichia coli.
OC      Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC      Escherichia.
OX      NCBI_TaxID=562;
[1]
RP      SEQUENCE FROM N.A.
RP      STRAIN=K12;
RX      MEDLINE=97251358; PubMed=9097040;
RA      Itoh T., Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,
RA      Kato H., Kimura S., Kitakawa M., Kitagawa M., Makino K., Miki T.,
RA      Mitsuuchi K. K., Mori H., Mori T., Motomura K., Nakade S., Nakamura Y.,
RA      Nishimoto H., Nishio Y., Oshima T., Saito N., Sampei G., Seki Y.,
RA      Sivasubraman S., Tagami H., Takeda J., Takemoto K., Wada C.,
RA      Yamamoto Y., Horikuchi T.;
RT      "A 460-kb DNA sequence of the Escherichia coli K-12 genome
RT      Corresponding to the 4.0-1.50.0 min Region on the Linkage Map.";
RL      DNA Rel. 3:379-392(1996).
RL      EMBL; D90836; BAA1579.1; -.
DR      InterPro: IPR003344; Big_1.
DR      InterPro: IPR003535; InlInlin.
DR      InterPro: IPR000601; PKD_domain.
DR      InterPro: IPR001639; T2SP_C.
DR      Pfam; PF02369; Big_1; 13.
DR      PRINTS; PR00810; BCTERIALGSPC.
DR      PRINTS; PR01369; INTMIN.
DR      SMART; SM00089; PKD_4.
NON_TER      2349      2349
SEQUENCE      2349 AA; 2349 MW; E45F5EE2AE2A76B59 CRC64;

```

alignment_scores:

Quality:	109.00	Length:	390
Ratio:	0.565	Gaps:	15
Percent Similarity:	49.487	Percent Identity:	20.000

alignment_block:

Align seg 1/1 to: P94750 from: 1 to: 2349

37 TCCTCAGTACTTTTCAATGTCGTATCGGCAGATAAANTCCCGAGA 86
746 SerAlaLysIleAlaIleuSerAlaSerAsnSngIYalLeuAlaAs 7620211aaS
87 TGAACCATPACTAAATATTGTTGGCCCCGCGAGACAGACGATCTTCCC 136
762 ngIuSnAlAlaIlaenTtValSerValAsnValAlaIaSpoluNlyser. 7780
137 CCAACATATATATTAAATTAACAATATTACAGCATAC..... AGT 1777

779AsnProIleasnAspHisThrValThlPheAlaValLeuSer	792
178	GAAAGTCATACCTGTGATGATAGAGTACTTTTATGTTGGCTCTCA	227
	: : : : : : : : :	:
793	GlySerAlaThrSerPheasnInsn	ThlAlaLy 805
228	CATTACACTTAATGGAGCATGTCCAACACAGTAGAATCCTAGACTTAT	277
	: : : : : : : : : :	:
805	sTrAspValasnGlyLeuAla...ThlPheaspLeuysSerSerLysG	821
278	CGGTGCAGCGGTGAACAATAATATACATTACATTTACGGAAMAAAAGA	327
821	IncluspasnThrValGlutValThlLeuGluasnGlyValylserInthr	837
	: : : : : : : : : :	:
328	TTPAATA.....	333
		:
838	LeuIleValSerPheValcLysPseSerThrAlaGlnValAspLeuCl	854
334AAAGAGACTACAAATTAAAGCCTATAAACAAATTA...TTGT	373
	: : : : :	:
854	nlySerLysasnGluValAlaValAlaasplYasnAspSerValThrmMet	871
374	TCAAAGTGTTTAACGCCCATCCGGCCCTAACACTTAATCAGCTCATATT	423
	: : : : : : : : : :	:
871	hpAlaThrValArspalArspalArspLysGlnLeuLeuAsnAspAlaMetVal	887
424	AACGTATAAATAAACCGCGCTCACGTGCACGTTATATATTATATATCC	473
	: : : : : : : : : : : :	:
888	ThlPheasnValasnSerAlaAlaValLysLeuSerGlnThlGluValAs	904
474	TGCTGGCGCACTAAAAAATTTGCTTTGGTGATCTGGAGTGCCTACTG	523
	: : : : :	:
904	nserHisAsp.....GlyLeuAlaThrAlaThrL	914
524	TGAGCTTAAGACTAAAAACAGATATAGCGAACCTAGCAACTTAACACT	573
		:
914	euthr.....SerLeuLysasnLyspyrArg	923
574	ATAAATATCATTTAAATTAATCATGATAGGAAGAAATATTCAATATGGTT	623
	: : : : : : : : : :	:
924	ValThlAlaSerValSerSerelGlyserGlnAlaasnGlnGlnValasnph	940
624	ACCCTGATTCAAAAGTAGACGCTCGCGTCGATCTTAACCTTGCCTCAAC	673
	: : : : : : : : : :	:
940	eileGlyaspGlnSerThrAlaAlaleuthrLeuSerVal...Proserg	956
674	GTGGGGCCACATATATATGGAAGAAATTCGTGATATGCTTTATGAT	723
		:
956	LyspIleThrVal.....	960
724	GGATATAGTACTAACGACGAGC.....TCCTTGGAGATAAGATV	761
	: : : : :	:
961ThlAsnThrAlaProGlnIntyrMetThrAlaThlLeuGlnAs	974
762	TCAGATTAACAATCCTAAATCTGATGGAAATTTATCTAAGAGAAATPA	811
	: : : : : : : : : :	:
974	pLysasnGlnYasnProLeuLysAspLyscLuiLeThrPheSerValProA	991
812	ATGATGTACCAACAAGAAATGCATATACTTTGTCACTTCTCTGGCGG	861
		:
991	snAsp.....ValAlaSerLysPheSerIleSerasnGlycyl	1003
862	AAAAGTTTAACTCCAACAATGAAGCAGCTGATTA.....	894
	: : : : :	:
1004	LysGlyMetThrAspSerasnLysAlaAlaIleAlaSerLeuthrGlyTh	1020
895AATATTG	901
		:
1020	rleuAlaGlyThrHisMetIleMetAlaArgLeuAlaasnSerAsnValS	1037
902	CTGACGACACTTCTCTGGAAACAACCTGGAATAGATTAACAGCTGTACC	951
	: : : : : : : : : :	:
1037	eTrAspAlaInProMetThrPheValAlaAspLysAspArgLysAlaVal	1053

```
952 ATGCCAGAAATCAGTGTCCGCTGTGTTGGCTGCA.....CGTTT 995
    :::  ::  |||  ::  |||  ::
1054 ValLeuGlnThrSerLysAlaGluIleIleGlyAsnGlyValAspGluTh 1070
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
996 GCATTGGATGCAAAAGTGAAAAATCCCGAGCTGGACAATATATGGGTA 1045
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1070 rThrLeuThrAlaThrValLysAspPro...SerAsnHisProValAlaG 1086
    ::|||::|||::|||::|||::|||::|||::|||::|||::|||
1046 ATATTATGTACTTTCACA 1065
1086 LylleThrValAsnPheThr 1092

seq_name: sp_bacteriap:P76347

seq_documentation_block:
ID   P76347.  PRELIMINARY:  PRT:  2383 AA.
AC   P76347.
DT   01-FEB-1997 (TREMBLrel. 02, Created)
DI   01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DE   01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE   02383.
GN   B1978.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RA   Blattner F.R., Plunkett G. III, Mayhew G.F., Perna N.T., Glasner F.D.;
RL   Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN   [2]
RP   SEQUENCE OF 2118-2383 FROM N.A.
RC   STRAIN=K12;
RA   Ikeda H., Baba T., Fujita K., Hayashi K., Honjo A., Horuchi T.,
RA   Ikemoto K., Inada T., Isono K., Isono S., Itoh T., Kanai K., Kasai H.,
RA   Kashiimoto K., Kim S., Kimura S., Kitagawa M., Kitakawa M., Makino K.,
RA   Masuda S., Miki T., Mizubuchi K., Mori H., Motomura K., Nakamura Y.,
RA   Nishimoto H., Nishio Y., Oshima T., Saito N., Sampei G., Seki Y.,
RA   Tagami H., Takemoto K., Wada C., Yamamoto Y., Yano M.;
RL   Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
DR   EMBL; AE000289; AAC75042.1; -.
DR   EMBL; D90837; BAA15800.1; -.
DR   InterPro: IPR003344; Big_1.
DR   InterPro: IPR003335; Intimin.
DR   InterPro: IPR002482; LysM.
DR   InterPro: IPR000601; PKD_domain.
DR   Pfam: PF02369; Big_1; 13.
DR   PRINTS: PR01369; INTIMIN.
DR   SMART: SM00257; LysM; 1.
DR   SMART: SM00089; PKD; 4.
KW   Complete proteome.
SQ
SEQUENCE 2383 AA; 251391 MW; 129975f4c64225c CRC64;
```

```
alignment_scores:
  Quality: 109.00      Length: 390
  Ratio: 0.565         Gaps: 15
  Percent Similarity: 49.487      Percent Identity: 20.000
```

alignment_block:

US-09-839-894-9 x P76347 ..

Align seg 1/1 to: P76347 from: 1 to: 2383

```
37 TCTTCACTGCTTTTACATTTGCTGATCGCGACATAAAATTCGCCGAGA 86
   |||:::  ::  |||:::  ::  |||:::  ::  |||:::  ::
762 SerAlaLysIleAlaThrLeuSerAlaSerAsnAsnGlyValLeuAlaAs 778
   |||:::  ::  |||:::  ::  |||:::  ::  |||:::  ::
87 TGAAGACATTAATATTTTGGCCCGCGTCGACAGACGAATCTTCCC 136
   |||:::  ::  |||:::  ::  |||:::  ::  |||:::  ::
778 nGluAsnAlaAlaAsnThrValSerValAsnValAlaAspGluGlySer. 794
```

```
137 CCAACATATATATTAAATACCATATTACAGCATAC.....AGT 177
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
795 .....AsnProIleAsnAspHisThrValThrPheAlaValLeuSer 808
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
178 GAAAGCATACCTCTGTATGATAGATGACTTTTATGTTGTCTTCCTCA 227
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
809 GlySerAlaThrSerPheAsnAsnGlnAsn.....ThrAlaIy 821
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
228 CAATACACTTAATGAGCATGTCACACCGATGAGATCCTAGACTTCAT 277
    ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
821 sThrAspValAsnGlyLeuAla...ThrPheAspLeuLysSerSerLysG 837
    ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
278 CGGTCAAGCGTGAACAAATATATACATTCAATTTACGAAAAAAGAACT 327
    ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
837 InclAspAsnThrValGluValThrLeuGluAsnGlyValLysGlnThr 853
    ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
328 TTAATA..... 333
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
854 LeuIleValSerPheValGlyAspSerThrAlaGlnValAspLeuG1 870
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
334 .....AAAGAGACTACAAATTAAAGCTTTAACAATTA...TTGT 373
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
870 nLysSerLysAsnGluValValAlaAspGlyAsnAspSerValThrMet 887
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
374 TCAAAAGTGTAACTGCCATCCCGGCTTAACACTTAATCAGCTCATTTT 423
    ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
887 hAlaThrValAlaGspAlaLysGlyAsnLeuLeuAsnAspValMetVal 903
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
424 AACTGTAAATAAAACGCGCTTCAGGTGCAAGTTTATTTATATATATTC 473
    ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
904 ThrPheAsnValAsnSerAlaGlnAlaLysLeuSerGlnThrGluVal 920
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
474 TCGTGGCGAAGCTAAAAATTTGCTTTGTGTATCTGGAGTCAATC 523
    ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
920 nSerHisAsp.....GlyIleAlaThrAlaThrL 930
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
524 TGAAGTTAAGATAAAAGACGATATAGTAGACCTATGGAACCTACACT 573
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
930 eutThr.....SerLeuLysAsnGlyAspTyrArg 939
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
574 ATAAATATCATTAATTAATTAATGATTAAGGAATATTCAGATATAGTT 623
    ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
940 ValThrAlaSerValSerSerGlySerGlnAlaAsnGlnValAsnPh 956
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
624 ACCTCATGTCAAAGTGCAGCTGCGCTCGATCTTAATCTGCGTCAAGC 673
    ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
956 eIleGlyAspGlnSerThrAlaAlaLeuThrLeuSerVal...ProSerG 972
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
674 GTGGGGGCACATATATTGGAAGAAATTCGTGATATGCTTTATGAT 723
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
972 LysAspIleThrVal..... 976
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
724 GGATATATGATTAACAGACAGC.....TCTTGGAGATAAGATT 761
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
977 .....ThrAsnThrAlaProGlnTyrMetThrAlaThrLeuGlnAs 990
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
762 TCAGGATTAACATTCCTAAATCTGATGGGAAATTTTCTAAGGAAATAA 811
    ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
990 pLysAsnGlnLysnProLeuLysAspLysGlnIleThrPheSerValPro 1007
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
812 ATGATGACACCAAGAAATTCATATACCTTGTCACTTCTTGGCGGGT 861
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1007 snAsp.....ValAlaSerLysPheSerIleSerAsnGlyGly 1019
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
862 AAAAGTTTAATCTCCACAAATGGAACGATCATTA..... 894
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1020 LysGlyMetThrAspSerAsnGlyValAlaIleAlaSerLeuThrGlyTh 1036
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
895 .....AATATTC 901
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1036 rLeuAlaGlyThrHisMetIleMetAlaArgLeuAlaAsnSerAsnValS 1053
```

902 CTGACGAGCTCTCTGGAACAACAACTGATAGATTACAGTGTGACC 951
:::|||||
1053 eRaSPaLaGInPromeThrPheValAlaSPaSPaRGAlaValVal 1069
:::|||||
952 ATGGCAAAATCAGTGTCCGGTGTGTGTGGCTGGA.....CGTTT 995
:::|||||
1070 ValLeuGInThrSerLysAlaGluIleIleGlyAsnGlyValaSPGIuTh 1086
:::|||||
996 GCAATGTGATGACAAAGTGGAAATCCCGAGCGACGAATATATGGGTA 1045
||| ||| |||
1086 rThrLeuThrAlaThrValLysaSPPro...SerAsnHisProValaIaG 1102
:::|||||
1102 LyIeThrValaSPnPhetHr 1108
:::|||||

seq_name: sp_bacteria:09RLA0

seq_documentation_block:
ID 09RLA0 PRELIMINARY; PRT; 872 AA.
AC 09RLA0:
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-OCT-2001 (TReMBLrel. 18, Last annotation update)
DE DNA POLYMERASE I.
GN POLA.
OS Rickettsia typhi.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_Taxid=785;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WILMINGTON:
RX MEDLINE=99416441; PubMed=10486973;
RA Anderson J.O., Anderson S.G.E.;
RT "Genome degradation is an ongoing process in Rickettsia."
RL Mol. Biol. Evol. 16:1178-1191(1999).
DR EMBL: AJ238757; CAB56108.1; -
DR HSSP: P05582; IKPD.
DR InterPro: IPR002421; 5_3_exonuclease.
DR InterPro: IPR002288; DNA_pol.
DR InterPro: IPR001098; DNA_pol_A.
DR InterPro: IPR000513; Exo_N.I.
DR InterPro: IPR003584; HHH.2.
DR Pfam: PF01367; 5_3_exonuclease; 1.
DR Pfam: PF02739; 5_3_exonuc_N; 1.
DR Pfam: PF00476; DNA_pol_A; 1.
DR PRINTS: PR00868; DNAPOLI.
DR SMART: SM00475; S3EXOC; 1.
DR SMART: SM00279; HHH2; 1.
DR SMART: SM00482; POLAC; 1.
DR PROSITE: PS00447; DNA_POLYMERASE_A; 1.
DR PROSITE: PS00482; POLAC; 1.
SQ SEQUENCE 872 AA; 99309 MW; BC2EBE3490879A08 CRC64;

alignment_scores:
Quality: 108.00 Length: 378
Ratio: 0.607 Gaps: 15
Percent Similarity: 47.090 Percent Identity: 20.370

alignment_block:
US-09-839-894-9 x 09RLA0 ..

Align seg 1/1 to: 09RLA0 from: 1 to: 872

4 AATAAGATTTTATTATTATTACATGTTTCTCTTCAGTACTTTTAC 53
|||||
383 AsnLysSerIleLysIleThrTySerLeuLysHisLeuLeuLysPh 399
|||||
54 ATTGGCTGATCGGACAGATAAATTCGCCGAGATGAAGCAT..... 96
|||||
399 eTyAlaAsnGlnSerHisIleThrAlaIleGluAspLeuLeuLeu 416
|||||

97ACTAATATTTTGGCCG 114
|||||
416 eTgInTyThrLeuSerAlaGlyLeuValGlnLysAsnLeuPheAla... 431
|||||
115 CGTGACAGGAACGAAATCTCCCCCAACATATATATTAATACCATAT 164
|||||
432GluThrLeuThrLysaSPAsnIleIleAsnGluSerAl 444
|||||
165 TACAGCATACGAGAAAGTCATCTGTATGATAGATAGATGACTTTTAT 214
|||||
444 aLysIleValIleAsnPhelIleSerLeuTyf...LysGlnThrLeuEua 460
|||||
215 GTTGTCTCTCACAAATACACTTAAT.....GGA 243
|||||
460 lAlaGlnLysAsnLysAlaPheArgLeuTyfArgGluIleAspLeuPro 476
|||||
244 GCATGTCCCAACCACTGAGAAATCTGACAGTTCATCGGTACGGTGAAC 293
|||||
477 ThrcySPheIleLeuAspLysMetGluLysIleGlyIleLysValaSPAl 493
|||||
294 AAATATACATTAATTAATTAAGGAAAAA.....AGAAGTTAATAAAA 337
|||||
493 aAspTyfLeuHisGlnLeuSerAspGluPheGlyThrcylLeuLysI 510
|||||
338 GAGAGCTACAAAT..... 351
|||||
510 lGluGlnGluIlePheAlaLeuSerGlyThrcylPheAsnIleAlaSer 526
|||||
352 ...AAGGCTATAAACAATTAATTTGTCAAAAGTGTAACTGCCATCCGG 398
|||||
527 GlnLysGlnLeuSerGluIleLeuPheLysMetGlnLeuProSerG 543
|||||
399 CCTAAGCTTAACACACCTCATTTTAAGTAAATTAATTAATTAATTAAT 448
|||||
543 yAsnThrLeuAlaLysThrSerTyfSerThLys..... 555
|||||
449 GTGCAAGTTTATTTATTTATTCCTGCTGCGCAATTAATTAATTTGCT 498
|||||
556AlaGlyIleLeuLysLysLeuSer 563
|||||
499 TTGCTGCTATCTGGATGCTACTCTGAAGTTAAGA..... 534
|||||
564 GluAspGlyTyfHisIleAlaThrLeuLeuLeuArgTyfArgGlnLeuTh 580
|||||
535GTAAGAAGAGATATAGTAGACCTATGAGACTTAACCTTAATA 580
|||||
580 rLysLeuLysAsnThrTyfThrasPserLeuProLysGlnIleAsnAs 597
|||||
581 TCACATTAATTAATTA.....ACTGATTAAGGA 606
|||||
597 lThrLysArgIleHisThrThrPheLysGlnThrSerThThrThGly 613
|||||
607 AATATTCAGATATGTTACTCACTCAATTAATTAATTAATTAATTAAT 656
|||||
614 ArgLeuSerSerGlnGluProAsnLeuGln.....AsnI 625
|||||
657 TAACTGCGTCCCACTGCTGGGGCACATATATTAATTAATTAATTAAT 706
|||||
625 eProThrasPserSerAspGly.....AsnLysIleA 636
|||||
707 ATATGCTCTTTAT.....GATGATAT.....AGTATTAACAGC 741
|||||
636 rGlnAlaIaPheIleAlaGluAspGlyTyfLysLeuIleSerAlaSPArg 652
|||||
742 AGCTCTTGGAGATTAATTT..... 762
|||||
653 SerGlnIleGluLeuArgIleLeuSerHisIleAlaAsnValaSPValle 669
|||||
763CAGATACAAATTCCTAATCTGATGGA 790
|||||
669 uLysGlnAlaPheIleAsnLysGluAspIleHisThrcIlnThraLacysG 686
|||||
791 AATTTTATCTAAGGAATAATATGATGACACCAAGAA..... 828
|||||

1062 CACACCA...AGTAGCAACA 1080
 ||||| :|||
 1323 ethrprolysglyserinthr 1330

seq_name: sp_fun1:096VN2

seq_documentation_block:

ID 096VN2 PRELIMINARY; PRT; 430 AA.

AC 096VN2; 01-DEC-2001 (TREMblrel. 19, Created)

DT 01-DEC-2001 (TREMblrel. 19, last sequence update)

DE EXTRACELULAR ALKALINE PROTEASE.

GN PEP.

OS Blumeria graminis.

OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Leotiomycetes;

OX Erysiphales; Erysiphaceae; Blumeria.

RN NCBI_TaxID=34373;

RP SEQUENCE FROM N.A.

RA Zhang Z., Guir S.J.;

RT "An extracellular alkaline protease from Blumeria graminis.";

RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AY039006; AAK84436.1; -

KM Protease.

SO SEQUENCE 430 AA; 44817 MW; 89B4D822BBBC3CB5 CRC64;

alignment_scores:

Quality: 105.50 Length: 397

Ratio: 0.603 Gaps: 21

Percent Similarity: 44.081 Percent Identity: 21.662

alignment_block:

US-09-839-894-9 x 096VN2 ..

Align seg 1/1 to: 096VN2 from: 1 to: 430

```

19 ATTTTTCATGTTGTTTCTCTCAGTACTTTTATACATTTGCTGATGGC 68
   :|||:|||||:|
12 ValIleThrLeuTyr.....LeuPheSerPheAlaTyrSerAl 24
   :|||:|||||:|
69 AGATAAATTT.....CCCGAGATGAAGATCACTAATATTT 106
   :|||:|||||:|
24 aProPheLeuPheSerAlaProAlaAsnProSerSerIleValAsnGlyL 41
   :|||:|||||:|
107 TTGGCCCGCGTGACGAGACGAATCTCCCAACAAATTAAT...ATAATTA 153
   :|||:|||||:|
41 euGlyLeuProIleSerAsnLeuAspAlaProAlaGluAspMetIleAla 57
   :|||:|||||:|
154 AAATACCATATTCACGATACAGT.....GA 179
   :|||:|||||:|
58 AsnArgTyrIleValIleValTyrSerLysAspAlaThrAspAlaValG 74
   :|||:|||||:|
180 AAGTCATACCTGTATGATGATGATGATGTTTATGTTGCTCTTCACA 229
   :|||:|||||:|
74 userHis.....GlnSerThrIleLysThrAlaLeuLysLysA 87
   :|||:|||||:|
230 ATACACTTAAT.....GAGACATGTCACCAACGATGAGAA 240
   :|||:|||||:|
87 rGSerLeuAsnAlaThrSerIleAspGlyArgGlnLeuSerPrometMet 103
   :|||:|||||:|
241 .....GAGACATGTCACCAACGATGAGAA 263
   :|||:|||||:|
104 HisThrPheLysMetGlyGlyTyrPArgIlyMetCysLeuAspAlaGluAs 120
   :|||:|||||:|
264 TCCAGACAGTTCATCGGTGACGCGTGAAACAAATATTAACATTAATTTA 313
   :|||:|||||:|
120 palameIleIleGluIleGlnSerAlaSerGluVal..... 132
   :|||:|||||:|
314 CGCAAAAGAAAGTTTAAATAAAAGAGAGCTACAAATTAAGGCTATATAA 363
   :|||:|||||:|

```

```

133 .....SerTyrValGluAlaAspThrGlnVal...GlyPheLeu 144
364 CAATTATTGTCAAAGGTTTAAGTCCCATCGGCCCAACA...CTTAA 410
   :|||:|||||:|
145 GluLeuThr...GluGlnIleGluAlaProIleGlyLeuValArgLeuSe 160
   :|||:|||||:|
411 CTCAGCTCATTTTAACGTGTATATAA.....AACGGCGCTT 445
   :|||:|||||:|
160 rHisAlaLeuLysSerGlnSerLysGluTyrValPheAspAsnAlaSerA 177
   :|||:|||||:|
446 CAGGTGCAAGTTTATATTATATAT...ATTCCTGCTGGCGAATAAAAAT 492
   :|||:|||||:|
177 spGlyAlaGlyIleValGlyTyrValIleAspThrGlnGlyIleArgAlaSer 193
   :|||:|||||:|
493 TTGCTTTTGGTGTATCTGGATGCTACTGTAAGTTAAAGTAAAAAG 542
   :|||:|||||:|
194 HisGluPheGlyGly..... 198
543 AGATATATAGTAGACACTATGAGACTTACACTATTAATATCACTATTAAT 592
   :|||:|||||:|
199 .....ArgA 200
593 TAACTGATTAAGGAATATTCAGATATGTTTACTCAGTTCAAAAGTGAC 642
   :|||:|||||:|
200 lArhMetLysAlaAsn.....PheIleAspAsp 209
643 GCTGCGCTCGATCTTAACCTGCGTCAACTGCTGGGGCAATATATTGG 692
   :|||:|||||:|
210 IleAsnGluAspGlnAsn.....GlyHisGlySerHisValAl 222
693 AAGAAATTCCTGTGATATATGCTTTTATGATGATATATAGTACTTAACGA 742
   :|||:|||||:|
222 aThrThrIleGlyGlyAlaSerPhe.....GlyValAlaLysAsnIleG 237
   :|||:|||||:|
743 GCTCTTGGAGATTAAGATTCAGGATTAACAATCCTAAATCGATGGGAAA 792
   :|||:|||||:|
237 luleuValGlyIleLysValLeuAspAlaLysGlyLysGlySerAsnAla 253
   :|||:|||||:|
793 TTTTATCTAAGAAATTAATGATGACACCAAGAATATGCTATACTTT 842
   :|||:|||||:|
254 AsnValLeuArgGlyValAsnPheAspValAlaGluArgGly..... 267
843 GTCACTCTCTTGGCGGTAAAGTTTAATCTCAACAANAATGAGACATAT 892
   :|||:|||||:|
268 .....LeuAlaGlyLysAlaVal.....ValAsnI 276
893 TAAATATTGCTGACGAGCTTCTCTGGAACAACATGGAATAGATTAACA 942
   :|||:|||||:|
276 lserIleGlyGlySerLysSerLysAlaLeuAsn...ThrAlaIleGlu 291
   :|||:|||||:|
943 GCTGTCACCATGCCAAGAATCAGTGTCCGCTGTGTGCTGGCTGGACG 992
   :|||:|||||:|
292 AlaleuThrLysAlaGlyValThrValValAlaAla..... 304
993 TTTCCAATTGGATGCACAAAGTGAAAAATCCGAGGCTGACATATATAGG 1042
   :|||:|||||:|
305 .....G 305
1043 GTAATATTAAAT.....GTACTTTCACACCAAGTAGT 1074
   :|||:|||||:|
305 LysAsnSerAsnLysAspAlaThrSerPheSerProAlaSer 318

```

seq_name: sp_bacteriap:09CFL1

seq_documentation_block:

ID 09CFL1 PRELIMINARY; PRT; 1441 AA.

AC 09CFL1; 01-JUN-2001 (TREMblrel. 17, Created)

DT 01-JUN-2001 (TREMblrel. 17, last sequence update)

DE UNKNOWN PROTEIN.

GN YOIC.

OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).

OC Ehrhartoideae; Oryzeae; Oryza.
 OX NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. NIPPONBARE;
 RA Buehl C.R., Yuan Q., Ouyang S., Moffat K.S., Hill J.N., Burr P.C.,
 RA Hsiao J., Zisman V., Pal G., Bowman C.L., Fujii C.Y., Vanaken S.E.,
 RA Bowman C.L., Craven B., Utterback T.R., Khalak H., Feldblum T.V.,
 RA Quackenbush J., White O., Salzberg S.L., Fraser C.M.,
 RT "Oryza sativa chromosome 3 BAC OSJNBa0040E01 genomic sequence."
 RT Submitted (Aug-2001) to the EMBL/GenBank/DBD databases.
 DR EMBL; AC079887; AAK16187.1;
 KW Hypothetical protein.
 SQ SEQUENCE 803 AA; 89209 MW; E119841096A8BDAC CRC64;

alignment_scores:
 Quality: 104.00 Length: 322
 Ratio: 0.717 Gaps: 10
 Percent Similarity: 45.031 Percent Identity: 20.186

alignment block:

US-09-839-894-9 x O9AUV5 ..

Align seg 1/1 to: O9AUV5 from: 1 to: 803

97 ACTAATATTTT...GGCCCGCGTGACAGACAGCATCTTCCCCAACA 143
 |||||
 406 ThrasnileuaspGlnSerSerLeuasnlysthrlyleSerH 422
 144 T.....AATATAT 151
 422 saspGlyleuValalaThrasnGlyValtrpGlylnHisArgAspValM 439
 152 TAAATAACAT...ATTACAGCATACAGTAAGTCACTACTGTATGAT 198
 :|||
 439 etGlnasnHsProleuGlyValtyrSerSerSerHsAlaArgHsPro 455
 199 AGCATGACTTTTATGTCTGTCTGCACATACACTTAATGAGCATG 248
 :|||
 456 GlnleuValleuPromethrCysThrSerHsValleuLeuProporPr 472
 249 TCCAACCCAGAGATCTAGACATTCATCGACGCGGTGAACAATA 298
 |||||
 472 oasnleuSerAsnHsProAsnSerSerPheValArgGlyThrIleuL 489
 299 TAACATTACAATTACGAAAAAAGATTAAATAAAGAGACATCAAA 348
 :|||
 489 yslvsAlaGlyleuMetGlyAlaTySerThrThrGluAsnTyLeuAsp 505
 349 ATTAAAGCCTATAACAATATATGTTTCAAAAGTGTAACTGCCATCCG 398
 :|||
 506 leuAsnHsSerGlyArg.....LysGlyLleIleCysProlyse 519
 399 CCTAACCTTAACCTCAGCTCATTTTAACTGTAATTAATAAAGCGGCTCAG 448
 :|||
 519 rserGlnHsValalaGluasnIleCysAsnlysalalaIuTyrg 536
 449 GTGCACCTTATATATATATTCCTGCGGCAACTA..... 486
 :|||
 536 IncysAsnAspTySerGlnIleValCysasnGlnInThrValleuL 552
 487AAAAATTGGCTTTGGTGATC.....TGGGA 515
 :|||
 553 ProleuasnlysalaserHsPheGlyLeuProthrlyslsPheAs 569
 516 TGCTACTCTGAAGTTAGAGTAAGAAACAGCATATAGTACGACATGGA 565
 :|||
 569 pGlylyleuValserArgGlnlyslsArglysalArgAlaThr..... 583
 566 CTTAACACTATAATATCATCTAATTAATTAAGTAAAGGAATATTTCA 615
 :|||
 584SerleuL 586

616 ATATGTTTACCTCAGTTCAAAGTAC...GCTGCGGCGATCTTAAGT 662
 |||
 587 SerThrPhtsalaglnValMetSerGlyCysSerlysthrHsIsthrAr 603
 :|||
 663 GCGTCCAACTGTGGGGGCACATATATTGAGAAATTCCTGTGATATGT 712
 :|||
 603 glyProGlnleuAspThrPhtHsAlaThrArgArgleuValGluLysV 620
 713 GCTTTTATGATGATATAGTACTACAGCAGCTTTTGGACATTAAGATT 762
 :|||
 620 alaPhtAlaGluAsnMetlysthrlyAsnSerThrPhe..... 632
 763 CAGGATTAACATCTTAATCTGATGGGAATTTATCTAAGAAATATAA 812
 :|||
 633Valse 634
 813 TGATGACACCAAGAAATTCGATATACCTTTGTCACTTTCGCGGATA 862
 :|||
 634 rGlnAlaGlnlysalArgleuAlaPheThrThrlyleuMetGlnTyrlle 651
 863 AAAGTTTAACTCCACAAATGGAACGTCATTAATATATCTGACGCACT 912
 :|||
 651 euProValleuProAspArgleuValalaAsnAlaIleAsp..... 665
 913 TCCTGGAACCAACTGGAATAGAAATTACAGCTGTCAACATGCCAGAA 962
 :|||
 666 SerGlyslutThrIleValTyArgThrSerArgleuAlaLeuProAspAl 682
 963 CAGTGTCCGCTGTG 978
 :|||
 682 aPheAsnProAlaIle 687

seq_name: sp_bacteriap.O9R2S7

seq_documentation_block:
 ID O9R2S7 PRELIMINARY; PRT: 691 AA.
 AC O9R2S7;
 DT 01-MAY-2000 (TReMBLrel. 13, Created)
 DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
 DE 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
 DE HYPOTHETICAL 69.7 KDA PROTEIN.
 GN DRB0037.
 OS Deinococcus radiodurans.
 OC Plasmid MPl.
 OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
 OX NCBI_TaxID=1299;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RI,
 RX MEDLINE=20036896; PubMed=10567266;
 RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
 RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
 RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
 RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zaleski C.,
 RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischman R.D.,
 RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
 RA Fraser C.M.;
 RT "Genome sequence of the radioresistant bacterium Deinococcus
 RT radiodurans RI.";
 RL Science 286:1571-1577 (1999).
 DR EMBL; AF001826; AAF12628.1; -.
 DR TIGR; DRB0037; -.
 DR InterPro; IPR002048; EF-hand.
 DR PROSITE; PS00018; EF_HAND; Gram_pos_anchor.
 DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
 KW Hypothetical protein; Plasmid; Complete proteome.
 SQ SEQUENCE 691 AA; 69702 MW; 744112A6A23C022E CRC64;

alignment_scores:
 Quality: 103.50 Length: 328


```
414 AGCTCATTTTAACTAATAATAAAACGGGCTCAGTGAAGTTAAT 463
      : ::::|:::
840 yIleTyRpheGly.....TyrL 846
      |||
464 TATATATTCCTGCTGGCAACTAAAAATTGGCTTTGGTGTATCTGG 513
      |||||
846 eutyRllePro.....LysAsnThrPro..SerGlyLeuTyr 857
      : ::::|:::
514 GATGCTACTCTGAAGTTAAGATAAAAGACATATAGTGAACCTATGG 563
      : ::::|:::
858 AsnValLeu.....PheAlaThrTyrTy 866
      |||
564 AACTACATATAATATCATTAATTAATTAAGGAAATATTC 613
      : ::::|:::
866 rSerTyThrLeuAsnThrThrIleArg..... 875
      : ::::|:::
614 AGATATGTTACCTCAGTTCAAAAGTGAGCGTCGCGATTTAACTTG 663
      : ::::|:::
875 ..... 875
664 CGTCCACAGTGGGGGACATATATTGAGAAATTCGTTGATATGTG 713
      || |||
876 .....GlyPheTyTyGlyGlnIle..... 882
      : ::::|:::
714 CTTTATGATGATATAGTACTACAGCAGCTCTTGAGATAGATTTC 763
      || : ::::|:::
883 .....TyrValSerAsnGlnAlaThrIleSerValLysServ 895
      : ::::|:::
764 AGCATAAACAATCTAATCTGATGGAATTTATCTAAGAAATAAAT 813
      : ::::|:::
895 alAsnTyRAlaPheGlnGlnThrValPheIleTyRAlaAsnIleThr 911
      : ::::|:::
814 GATGACACCAAGAAATGCAATTAATCTTGTCACTTCTTGCGGGGTAA 863
      : ::::|:::
912 AsnGlyThrAsnGlnIleLysPhe..... 919
      : ::::|:::
864 AAGTTAACTCCACAAATGGAAGACGTCATTAATATGCGACGACGTT 913
      |||
920 .....GlyMetPheSerAlaThrValTyrProSerS 930
      : ::::|:::
914 CTCGGAACAACAATGGAATAGATTACAGCTGCACCATGCCAGAAATC 963
      |||||
930 erLeuSerPheAsnTyThrThrIleSerSerIle.....Ile 942
      : ::::|:::
964 AGTGTTCGGGTGTGTGTGGCCTGAGACGTTTGCAATGAGATCAAAAGT 1013
      : ::::|:::
943 GlnIleProLeu.....Trp..... 947
      : ::::|:::
1014 GGAATAATCCGAGAGCTGACAAATATATGGTATATTAATGTTACTTCA 1063
      |||||
948 .TyrAsnProLysIleGlnIleTrpGlnGlyAsnPheThrLeuProSerA 964
      : ::::|:::
1064 CACCAAGTACTCAACACATC 1083
      : ::::|:::
964 lalIleSerAlaGlnGlyAsnLeu 970
```

7

This Page Blank (uspto)